The Dissertation Committee for Kevin Robert Burnham Certifies that this is the approved version of the following dissertation:

Phonetic Training For Learners of Arabic

Committee:

Mahmoud Al-Batal, Supervisor

David Birdsong

Kristen Brustad

Mohammad Mohammad

Harvey Sussman

Phonetic Training For Learners of Arabic

by

Kevin Robert Burnham, B.A.; M.A.; M.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August, 2013

Dedication

To those poor Ephraimites

Acknowledgements

Space scarcely permits to thank all of the people that helped make this dissertation possible. I would like to thank Dr. Mohammad Mohammad for agreeing to join the committee at the last minute and Dr. Martha Schulte-Nafeh for her earlier assistance with the dissertation. Thanks also to Dr. Harvey Sussman for his advice and suggestions which helped to refine the final product. I owe my gratitude to David Birdsong for teaching me how to do empirical research and for making sure that this dissertation was properly designed and implemented. And I must give special thanks to Dr. Mahmoud Al-Batal for his immeasurable support of every kind since my arrival in Austin nearly 6 years ago and especially in helping to make sure that this dissertation was completed successfully.

This experiment also would not have been possible without the assistance of Corinne Stokes who ran a large number of participants at UT Austin during early phases of the research when I was not able to be on campus. Her attention to detail and tireless efforts are well appreciated. I must also thank Dr. Rajka Smiljanic and other members of the UT Sound Lab for kindly allowing us the use of that lab. The process of recording stimuli for this dissertation was greatly facilitated by their generosity.

I would like also to thank my colleagues at the other universities that were part of this experience, especially Summer Loomis and Peter Glanville, but also to all those teachers who made their students aware of my research, and encouraged them to participate. I hope that the results of this experiment are valuable to them.

I thank also Kirsten Comandich whose professionalism, programming skills and enthusiasm for the research helped see it through to its successful conclusion.

Thanks also to my native speaker informants and all those that gave their time to participate in this experiment. It is heartening to see that so many are willing to donate their time to help others to learn Arabic.

This research could not have been completed without a generous contribution from the Arabic Flagship program. I hope sincerely that its students will be able to reap the rewards now and in the future.

Last but not least I thank the friends and family that supported me unfailingly throughout this long process. My parents, whose support has never flagged, my sisters who can always cheer me up, and my beloved wife Anousha whose patience has been biblical.

Phonetic Training For Learners of Arabic

Kevin Robert Burnham, Ph.D.

The University of Texas at Austin, 2013

Supervisor: Mahmoud Al-Batal

This dissertation assesses a new technique intended to improve Arabic learning outcomes by enhancing the ability of learners to perceive a phoneme contrast in Arabic that is notoriously difficult for native speakers of English. Adopting a process approach to foreign language listening comprehension pedagogy, we identify and isolate an important listening subskill, phonemic identification, and develop a methodology for improving that skill. An online training system is implemented that is based upon known principles of speech perception and second language speech learning and has previously been used to improve phonemic perception in a laboratory setting.

An empirical study investigating the efficacy of the training methodology was conducted with 24 2nd and 3rd year students of Arabic in several different intensive Arabic programs in American universities. The contrast under investigation was the Arabic pharyngeal (/ħ/) versus laryngeal (/h/) voiceless fricatives. Training participants completed 100 training modules, each consisting of a 24 item minimal pair test featuring the /ħ/-/h/ contrast in word initial position for a total of 2400 training trials over 4 weeks. The training website design was based on the high variability training protocol (Logan, Lively & Pisoni, 1991). The experiment finds significantly greater improvement ($F_{1,22}$=8.89, p = .007, η_2 = .288) on a minimal pair test contrasting /ħ/ and /h/ for a group

that received approximately 5 hours of phonetic training (n=10) compared to a control group (n=14) with no training.. Critically, these perceptual improvements were measured with stimuli that were not part of the training set, suggesting language learning and not just stimulus learning. Qualitative data from participants suggested that these perceptual gains were not restricted to the simple minimal pair task, but carried over to listening activities and perhaps even pronunciation.

The dissertation concludes with a discussion of phonemic perception and foreign language instruction and implementation of phonetic training within an Arabic curriculum.

Table of Contents

List of Tables

List of Figures

Chapter 1: Listening Comprehension Pedagogy

1.1 INTRODUCTION

This dissertation represents an effort to contribute to Arabic language pedagogy by improving the ability of learners of Arabic to accurately perceive the phonemes of that language. The theoretical foundations of the research are located at the intersection of three fields of inquiry. The first of these is the noted difficulties that native speakers (NS) of English have in perceiving and pronouncing a number of different Arabic phonemes, most notably the guttural sounds that are produced with articulations deep in the throat. The second is what has been termed the process approach to listening comprehension pedagogy which advocates breaking down the components of listening comprehension into its various sub-skills and designing specific activities that are intended to improve those skills in language learners. Finally, the dissertation draws upon findings in the field of second language speech perception, specifically phonetic training, to design and test a website that trains learners of Arabic to more accurately perceive the most difficult Arabic phoneme the pharyngeal fricative /ħ/. The purpose of this introductory chapter is to first more precisely define the specific pedagogy problem that this dissertation intends to address and to examine how the problem is currently approached by Arabic language pedagogy in the United Stated. It will then examine the specific articulatory and acoustic correlates of the pharyngeal /ħ/ in comparison with the laryngeal /h/ with which it is frequently confused. Following that discussion, the chapter will discuss the process approach to listening comprehension which provides the theoretical justification for the research within the field of second language pedagogy. As shall be seen in this discussion, the process approach advocates a more in-depth understanding of the physiological processes that contribute to successful listening so that listening sub-skills can be isolated and targeted for improvement. It will be the purpose of the next three chapters of this dissertation to provide this in-depth understanding.

In Chapter 2 the process of phonemic perception in native speakers is examined so that a clear picture of the deceptively simple process of identifying a given phoneme based on an acoustic signal is better understood. This chapter lays the groundwork for Chapter 3 which examines speech perception in second language learners and the challenges they face in adapting their native language speech perceptual system to accurately categorize the novel set of speech contrasts in their second language. Finally, Chapter 4 presents an in depth investigation of phonetic training, a laboratory based methodology that has been shown to improve phonemic perception in at least some tasks after mere hours of training, even for learners whose perceptual system failed to make the necessary adjustments after long exposure to their second language. The purpose of this chapter is to elucidate the best practices of phonetic training so that an effective training program can be designed for learners of Arabic that wish to improve their ability to perceive the Arabic /ħ/.

Chapter 5 describes the results of an empirical experiment conducted with a group of Arabic learners in which the effectiveness of the training website is assessed and also presents qualitative data obtained from participants in the phonetic training program. Finally, Chapter 6 discusses the prospects for phonetic training as a pedagogical tool in Arabic and beyond and lays out a vision for how the website could be improved, such that training is more effective, and expanded such that other phonemic contrasts could be incorporated. The research presented here is only the first step in the planned development of a suite of phonetic training modules that diagnoses individual problems, designs a training regimen specific to individual learners and tracks training progress over time.

1.2 TEACHING ARABIC LANGUAGE PHONOLOGY

It is difficult to generalize at present as to how Arabic phonology is taught across a variety of programs at universities in the United States, and there appears to be no published research or articles that focus on this issue. Therefore, only tentative conclusions may be drawn here. The author of this dissertation has experience as a teacher at several of the leading Arabic teaching institutions across the United States and has experience as a student and teacher with the leading textbooks used in teaching Arabic. Furthermore, he has discussed many of these issues with colleagues in the field and conducted some preliminary research. The description of how Arabic phonology is taught cannot, of course, apply to all teachers and all programs at all times.

According to a 2009 survey (Abdalla & Al-Batal, 2012) the *Alif Baa/Al-Kitaab fii Ta^{c}allum al-^{c}Arabiyya* series first published in 1995 and now in third edition by Brustad, Al-Batal and Al-Tonsi is currently being used by 90% of the Arabic teachers at the college level in the United States who responded to the survey[1]. The *Alif Baa* book is designed for teaching the Arabic writing and sound system as well as basic vocabulary and phrases. In its most recent incarnation, the book is accompanied by a website which provides some of the listening drills online. In teaching students to perceive and pronounce the novel phonemes of the Arabic language the book provides short acoustic and sometimes articulatory descriptions of the sounds along with audio examples found either on the website or on an accompanying DVD. Tests of perception are provided

[1] There were 209 teachers that completed the survey representing approximately half of all Arabic teachers in the United States

either in the form of dictation exercises in which students listen to words (and also see video of an individual pronouncing the words) and are asked to write those words down. These assignments are then submitted to an instructor for grading and feedback. For particularly hard to perceive phonemes there are minimal pair tests in which students hear a word and must choose which word they heard from a limited set. For these activities immediate feedback is provided by the website. Upon completion of the *Alif Baa* book, learners transfer to the *Al-Kitaab* which transitions to a more overt focus on proficiency in all four language skills as well as providing opportunities for students to learn about Arab culture, including both little "c" culture (everyday customs and behaviors) and big "c" culture (knowledge about important persons, historical dates, works of art etc.). Phonologically, this book provides students with further opportunities to develop both perceptual and productive abilities. Each chapter provides a series of sentences presented in audio form only that students are asked to transcribe, thus providing further practice at establishing the correspondence between spoken and written forms of the language. However, since the majority of vocabulary found in these sentences is presumed to be known by the learners, they can rely on their knowledge of that vocabulary to disambiguate otherwise similar sounding phonemes and so these exercise cannot be said to focus on segmental level perception. For production each chapter comes with a short reading passage that students are expected to read aloud, record, and submit to their teacher for feedback. From a perspective informed by research in second language speech perception (to be reviewed more thoroughly in Chapter 3) the *Alif Baa/Al-Kitaab* series has a number of features to recommend it. For one, the early stages of sound learning present learners with both audio and video of individuals pronouncing these sounds. The visual modality has been shown to be an important contributor in helping learners to perceive and pronounce new sounds. Furthermore, the fact that learners are exposed to several different voices over the course of their sound learning is important. More diverse sound sets contribute to more generalized learning. Finally, the use of immediate feedback for learners, on at least some occasions, is important for helping learners to focus in on the relevant acoustic characteristics of Arabic sounds. These features make, the *Alif Baa/Al-Kitaab* series valuable in terms of its contribution to learners' working phonological knowledge of Arabic. However, these books alone are unlikely to be sufficient for all learners and at least some of them are likely to require additional attention and significantly more training than provided by the *Alif Baa* website.

It is of course more difficult to generalize about the behavior of individual teachers in the classroom in regards to the pronunciation and perception of Arabic phonemes. Nevertheless, anecdotal evidence and personal experience seem to suggest that phonology is deprioritized in the Arabic classroom when compared with vocabulary, grammar and cultural knowledge. This impression is reinforced by the paucity of published research on classroom phonology instruction. A noted exception is the study by Huthaily (2008) who describes an experiment he conducted in attempting to improve the perception and production of Arabic phonemes in his students. His methodology included teaching students to write in the IPA (International Phonetic Alphabet) as well as detailed descriptions of articulators and their actions in producing Arabic phonemes. Huthaily concludes that his training group showed significantly greater gains in both perception and production for at least some phonemes in comparison with an untrained control group that used an earlier edition of the *Alif Baa* book described above that did not have an accompanying website. However, Huthaily makes no effort to consider the demands placed on teachers in implementing his methodology. It seems certain that teaching students the IPA, and about articulators, will take time away from the important task of developing communicative fluency in learners. Furthermore, applying this approach with all students is unnecessary since some are clearly able to hear and produce the sounds of Arabic without specific intervention. Furthermore, as Huthaily himself notes, implementation of this methodology requires that teachers themselves be experts in phonology. Issues with phonology tend to vary substantially from one individual to another and it would be difficult to design classroom activities appropriate for all students in a class. Therefore, the phonetic training methodology to be described later in this dissertation, which can be targeted specifically at those individuals that need it, represents a potential solution to persistent perceptual and productive difficulties found even in advanced students of Arabic and can do so without making demands on class time and without requiring students or teachers to spend time and energy acquiring complex metalinguistic knowledge

1.3 Phonemic Perception in Learners of Arabic

Just as there is little research on how Arabic learners are taught phonology, there are few published studies that attempt to measure the learners' ability to accurately perceive and pronounce Arabic phonemes.

A notable exception is a 1982 study by Asfoor. In examining the production of Arabic consonants Asfoor first surveyed teachers at the Defense Language Institute in Monterrey, CA to find out which consonants seemed most difficult for NS of English. Based on these interviews she chose the ten most difficult (*ʕayn* /ʕ/, *ɣayn* /ʁ/, ħaːʔ /ħ/, *χaːʔ* /χ/, *dˤaːd* /dˤ/, *sˤaːd* /sˤ/, *qaːf* /q/, *hamza* /ʔ/, *tˤaːʔ* /tˤ/, and *ðˤaːʔ* /ðˤ/)[2] and elicited samples of each from Arabic learners both before and after an intensive 6-week language course. The recordings were analyzed by native speakers (NS) of Arabic and assigned

[2] /ʕ/ = voiced pharyngeal fricative, /ʁ/ = voiced uvular fricative, /ħ/ = voiceless pharyngeal fricative, /χ/ = voiceless uvular fricative, /dˤ/ = pharyngealized voiced alveolar plosive, /sˤ/ = pharyngealized voiceless alveolar fricative, /q/ = voiceless alveolar plosive, /ʔ/ = voiceless glottal plosive, /tˤ/ = pharyngealized voiceless alveolar plosive, /ðˤ/ = pharyngealized voiced dental fricative)[2]

acceptability ratings. Asfoor found that the ʕ*ayn* in particular was not only poorly produced, but that it seemed to be rather resistant to improvement, unlike several of the other phonemes. She also identified the emphatic *tˤa:ʔ* as an area in need of focus. Asfoor's data also indicate that the ħ*a:ʔ* showed little improvement such that it was the second most difficult to produce after 6 weeks. Asfoor's experiments were not concerned with segmental perception, so no measures of perceptual ability were taken.

Burnham (2010) tested NS of English learning Arabic with varying levels of Arabic experience with a minimal pair test featuring seven contrasts that were chosen based on interviews with other teachers of Arabic. The seven pairs are shown here:

Letter names	IPA symbols	Articulatory Descriptions
si:n – sˤa:d	/s/ - /sˀ/	Voiceless alveolar fricative vs. pharyngealized voiceless alveolar fricative
ta:ʔ - tˤa:ʔ	/t/ - /tˤ/	Voiceless alveolar stop vs. pharyngealized voiceless alveolar stop
ha:ʔ - ħa:ʔ	/h/ - /ħ/	Voiceless glottal fricative versus voiceless pharyngeal fricative
ɣayn – χa:ʔ	/ɣ/ - /χ/	Voiced velar fricative versus voiceless velar fricative
da:l – dˤa:d	/d/ - /dˀ/	Voiced alveolar stop versus pharyngealized voiced alveolar stop
ka:f – qa:f	/k/ - /q/	Voiceless velar stop versus voiceless uvular stop
ʕayn – hamza	/ʕ/ - /ʔ/	Voiceless glottal stop versus voiced pharyngeal fricative

Table 1.1: Difficult Arabic discrimination for NS of English

A minimal pair test was given to 8 NS of Arabic and 49 NS of English learning Arabic who were divided into 4 groups according to their current year of Arabic study (1^{st}, 2^{nd}, 3^{rd}, 4^{th} or beyond).

The experiment revealed clear differences in the learnability of the different pairs. For example, the /ɣ/-/x/ pairs were reliably discriminated at native-like levels by many of the subjects, including several from the 1st year group. In the advanced group all but a few subjects tested within NS range. The /ʕ/-/ʔ/ contrast showed somewhat poor performance at lower levels, but many learners in the advanced group were within NS range. Thus while the contrast is difficult at first, it does appear to typically be learned over time. On the other hand, results for /s/-/sˤ/ and /h/-/ħ/ showed very slow learning of the contrast. For both of these contrasts even the advanced learners, as a group, did significantly worse than the NS. There was only one instance of native-like perception of either contrast. Indeed many in the advanced group performed little better than 50% correct on the test, suggesting that they were just guessing. Furthermore, although both contrasts seemed to show a learning trend, there was no statistically significant difference between the beginning and advanced groups. The results from this study provide further empirical evidence of difficulties for nonnative speakers in perceiving L2 speech sounds and confirm that at least some students of Arabic fail to learn to perceive some contrasts in that language even after many hours of training. Unlike Asfoor, Burnham found evidence that perception of the *ʕayn* was well learned and most of the advanced group in that study were close to 100% correct on the task. This may be due to differences in the experimental design (Asfoor's was a longitudinal study) or the fact that the learners in Asfoor's study had much less experience with Arabic than the advanced group in Burnham's study. However, it would not be surprising to find that the *ʕayn* is more easily perceived than pronounced for NS of English. The articulation for the *ʕayn* involves a pharyngeal constriction (see details below) that is both novel and difficult for learners such that it is not uncommon for novice learners of Arabic to complain of muscle soreness when initially practicing this articulation. It also may reflect the fact that the task in Burnham's experiment was a simple one and allowed subjects to focus their full attention on perception alone. Anecdotal evidence suggests that at least some advanced learners continue to have difficulty perceiving this sound while performing higher order tasks like listening comprehension. Burnham also found evidence that the *χa:ʔ* /χ/, *ɣayn* /ɣ/ and *qa:f* /q/ were well learned by his sample.

In addition to these empirical results, Burnham(2010) interviewed 7 teachers of Arabic and asked them which phonemes, in their experience, were the most difficult to perceive for their learners. All seven of these teachers indicated that the /ħ/-/h/ discrimination was especially difficult for learners of Arabic and most also mentioned the emphatic consonants. Furthermore, one veteran teacher with many years of experience teaching at all levels, has indicated that the /ħ/-/h/ contrast is the most difficult for learners and even some who have achieved superior proficiency continue to struggle with it.

Based on the above studies the /ħ/-/h/ contrast was selected as the target of phonetic training for this dissertation. It seems that training with this contrast offers the greatest opportunity for improved outcomes in Arabic pedagogy. Furthermore, if perception of this extremely difficult contrast can be improved with phonetic training it seems reasonable to conclude that other difficult contrasts found in Arabic could as well. Thus the /ħ/-/h/ contrast provides a strong test of phonetic training as a pedagogical tool.

1.4 ACOUSTIC AND ARTICULATORY CORRELATES OF THE PHARYNGEAL LARYNGEAL CONTRAST

Laboratory studies of the Arabic /ħ/, usually described as a voiceless pharyngeal fricative, have been primarily focused on identifying the specific articulation that is used to produce the sound, and have generally compared it to the Arabic /ʕ/ (*ʕayn*), which is included on the IPA chart as a voiced pharyngeal fricative (IPA Chart, 2005). Most of the research in this area has been motivated by a phonological question, mainly why do all of the Arabic gutturals – uvulars, pharyngeals and laryngeals – appear to pattern together phonologically when they do not share a common articulation (see McCarthy, 1994). Studies of the acoustic characteristics of /ħ/ are more rare, and there is only one published study that explicitly investigates the acoustic cues that NS of Arabic use in discriminating between /h/ and /ħ/. Not all of the studies below specifically compare /h/ with /ħ/, however, they all attempt to describe at least one of those two phonemes articulatory, acoustically, or both. These studies are included here because they contribute to our understanding of how the pharyngeal /ħ/ is articulated, and what it sounds like

Obrecht (1968) provides acoustic descriptions of both /h/ and /ħ/ based on data from a single Lebanese subject. For /ħ/ he notes that it is characterized by aperiodic noise down to about 1750Hz "with the areas of greater intensity anticipating F2 and F3" (26) and the strongest noise area between 2100 and 3600Hz. He states that /ħ/ is "in effect a replacement for voice" (26). He describes /h/ relative to /ħ/ and claims it is "lighter in intensity and somewhat higher in frequency" (27). He says that while /ħ/ fills F3 and F2 with traces in F1, /h/ fills F3 and F4 with traces in F2 and that furthermore /h/, (75-100ms), was shorter than /ħ/ (150ms).

Delattre (1971) defines a pharyngeal articulation as "one in which the root of the tongue assumes the shape of a bulge as is drawn back toward the vertical back wall of the pharynx to form a stricture." (129). This divides the vocal tract into two distinct cavities, one beneath the stricture and extending to the glottis and the other above it and extending to the lips. He uses Arabic as a prototypical case and compares its pharyngeals with pharyngeal articulations in German, Spanish, French and English. He claims 5 different pharyngeals for Arabic, adding /ɣ/, /χ/ and /q/. He made x-ray films from a single subject who was described as an educated native speaker of Lebanese. He elicited the tokens in word initial position followed by three different vowels /i/, /a/ and /u/. In comparing /ʕ/ and /ħ/ he claims that the two are a voiced/voiceless pair and that the pharyngeal stricture for /ħ/ is lower and narrower than that of /ʕ/ and that therefore the lower cavity is smaller in producing the /ħ/. Acoustically, he notes that the first formant transition for /ħ/ is characterized by aperiodic noise and is slightly higher than that of /ʕ/. Delattre concludes that /ħ/ should be considered an emphatic counterpart to /χ/ and /ʕ/ an emphatic counterpart of /ɣ/; however, this conclusion does not seem warranted as there is no evidence of a significant drop in F2 as is found for vowels in the area of other emphatic consonants (Jongman, Herd, Al-Masri, Serono & Combest, 2011). Finally Delattre claims that these sounds are not significantly affected by the sounds around them.

Ghazeli (1977), a NS of Tunisian Arabic, used cineflouregraphic film of his own productions of a number of words featuring the two pharyngeal consonants /ħ/ and /ʕ/ and also gathered acoustic data from 12 other subjects from 6 different Arab countries. All subjects were adult males and included 6 NS of Tunisian Arabic, 2 of Libyan, 1 Algerian (Algiers), 1 Egyptian (Cairo), 1 Northern Bedouin Jordanian Arabic, and 1 Iraqi (Baghdad). He found that for both pharyngeals there was a contraction of the laryngopharynx as a result of the backward movement of the tongue and forward displacement of the back wall of the pharynx. The constriction caused by these articulations was very narrow, approximately 3mm for /ħ/ and 4mm for /ʕ/. The author surmises that the greater constriction for /ħ/ may be due to the fact that it is not accompanied by vocal fold vibration and so requires a narrower constriction for the gesture to produce audible sound. He also took measurements of nasal airflow and found no evidence for the nasalization of these pharyngeals contrary to results reported by Delattre (1971).

Acoustically, Ghazeli notes a strong influence of neighboring vowels on pharyngeals. He gives the example of the syllable /ʕi/ in which the vocal tract takes the shape necessary for /i/ even before the voicing for /ʕ/ begins so that the transition merely involves a widening of the laryngopharynx and narrowing of the oral cavity. The /ħ/ is described acoustically as having "strong aperiodic noise but with visible formant structure "(Ghazeli, 1977:45). Like /ʕ/ the formant structure of /ħ/ is similar to that of adjacent vowels and like Delattre (1971) he found that the pharyngeal constriction for /ħ/ was more extreme than that of /ʕ/. Ghazeli states that these pharyngeals have no coarticulatory effects on surrounding consonants and only minor effects on adjacent vowels, most notably a slight rise in F1. It is not clear though what Ghazeli is comparing these effects to. He notes that these effects varied from one dialect to another with Tunisian, Algerian, Libyan and Egyptian subjects showing the most sizeable change in F1 while the Iraqi and Jordanian subjects showed little change. Of course, with the small number of subjects it is unclear if these are interdialectical or interspeaker differences.

Laufer & Baer (1988) examined both primary and secondary pharyngeal articulations in NS of Hebrew and Arabic. Their subjects included 5 NS of Hebrew, three of whom also spoke Arabic, and 4 NS of Arabic. They inserted a fiberoptic endoscope through the nasal passages of the subjects to observe the action of the pharynx while they elicited pharyngeals articulations in real words, nonsense words and sentences. They indicate that all of the subjects used similar articulatory strategies for producing pharyngeals. The primary conclusion of the study is that the primary place of articulation for /ħ/ and /ʕ/ is identical to the secondary place of articulation for the pharyngealized consonants described above.

Yeou & Maeda (2011) also compare the pharyngeals /ħ/ and /ʔ/ with the uvulars /χ/ and /ɣ/. They develop an idealized model of vocal tract configurations for both pharyngeals and uvulars and compare its predictions with the productions by 4 male NS of Morocaan Arabic. As predicted by their model, the first two formants in the pharyngeals is higher than that of uvulars while the third formant is lower. Their spectrograms of /ħ/ "shows zones of frication noise characterized by formant structure corresponding to F1, F2, F3 and F4." They found the F2 and F3 were always excited by the noise source while F1 and F4 varied considerably in the extent to which frication noise was noticeable. They conduct a second experiment to determine if these phonemes are best described as fricatives or approximants by comparing /χ/ and /ħ/ to the known fricative /s/. They elicited speech samples using real word tokens featuring the phones in initial, intervocalic and final position. They find that the amount and pattern of airflow for /χ/ and /ħ/ is significantly different than that of /s/ and conclude that these two phonemes are best classified as approximants per the definition of Catford (1977 in Yeou & Maeda, 2011) in which turbulent airflow may accompany a voiceless approximant. These results are in line with those of both Ghazeli (1977) and Butcher & Ahmad (1987).

Heselwood & Al-Tamimi (2011) directly compare the articulatory and acoustic correlates of the pharyngeals /ħ/ and /ʕ/ with the laryngeals /h/ and /ʔ/. Their subjects were 7 NS of Jordanian Arabic (3 females) from several different cities. Some of the subjects provided nasoendoscopic data, some videofluoroscopic data and all but one provided acoustic data. Subjects read real words tokens featuring these phonemes in /a_a/ and /i_C/ contexts. The nasoendoscopic data revealed that in the production of pharyngeals the epiglottis shows "consistent and considerable retraction and lowering" (111) while for the laryngeals there was minimal retraction and very little lowering. They also found that the lateral pharyngeal walls advanced inward for some of the pharyngeals. The general pattern shown by the videofluoroscopic analysis was a much smaller pharyngeal area during the production of the pharyngeals "with the root of the tongue and the epiglottis much closer to the rear pharyngeal wall" (113). Although the hyoid bone was raised in the pharyngeal /ʕ/ compared to the laryngeal /ʔ/, this was not evident in comparing /ħ/ with /h/.

The acoustic analysis focused on the onset of F1 and F2 in the following vowels. They measure the values using the Bark scale. The Bark scale is a logarithmic scale that reflects the fact that the human auditory system is more sensitive to frequency differences at lower frequencies than at higher ones such that a change at higher frequencies must be greater in absolute terms to be noticeable in comparison to changes at lower frequencies. The Bark scale adjusts for this such that a difference of 1 on the Bark scale is equally detectable at all frequencies. The experimenters state that when F1 and F2 are within 3.5 Z on this scale they are integrated into a single perceptual formant and that the closer they are, the more they will be integrated. In analyzing the acoustic data provided by the subjects they find that F1 and F2 in vowels following laryngeals were almost always greater than 3.5 Z while those of pharyngeals were almost always less. They suggest that this is the perceptual cue that NS of Arabic rely on in discriminating between /h/ and /ħ/. Unfortunately, there do not appear to be any studies that test this supposition. In their conclusion, the authors argue that the pharyngeals can rightly be regarded as emphatic laryngeals so that the phonological term 'emphatic' can group all phonemes with a pharyngeal articulation into a single class.

Pharyngeal and laryngeal articulations are also compared by Shahin (2011) who gathered laryngoscopic data from a single Palestinian subject producing real word tokens with the target phonemes in a /ə_iː/ context (except for /ʕ/ which was word initial). She reports no pharyngeal constriction for the laryngeals and no retraction of the tongue root or epiglottis as found with the pharyngeals. Furthermore, the /h/ was realized as a voiced [ɦ] intervocalically while the spectrograms provided suggest that the /ħ/ remained voiceless between vowels.

In examining the evidence available, Ladefoged & Maddieson (1996) conclude that the two sounds we have been referring to as pharyngeal fricatives (/ħ/ and /ʕ/) are "neither pharyngeals nor fricatives" (168). Instead they consider them to be epiglottal approximants and should thus be represented by /ʜ/ and /ʢ/. However, we will retain /ħ/ and /ʕ/ as these symbols are used in all recent studies of Arabic phonetics.

Butcher & Ahmad (1987) elicited nonsense word tokens from 3 NS of Iraqi Arabic contrasting the two pharyngeals /ħ/ and /ʕ/ with two glottals /h/ and /ʔ/ in word initial and word final position in a variety of vocalic environments. They analyzed the tokens in spectrographs, but also with an oscillograph (to measure the duration and intensity of frication noise), and a pneumatograph (to measure airflow volume). They found that while all of the tokens of /ħ/ were accompanied by voiceless frication, none of those of /ʕ/ were. They also claim that the airflow rate for /ħ/ was higher than normal for fricatives, but lower than for /ʕ/ which they subsequently classify as a voiced approximant. In comparing the two pharyngeals to the glottals they note that the former cause a substantial rise in F1 in both following and preceding vowels. Also, with back vowels the pharyngeals caused a rise in F2 and with front vowels a drop in F2 compared to the pharyngeals. They conclude that "The phoneme /ħ/ is consistently realized as a voiceless continuant sound with high rates of airflow, high intensity and marked formant structure." (Butcher & Ahmad, 1987:170). The /ʕ/ on the other hand is classified as a voiced approximant with evidence of creaky voice caused by laryngeal constriction. They believe that the vocal tract configuration of for these two sounds is more or less equivalent.

There appears to be only one study that attempts to determine which acoustic cues are relied upon by NS of Arabic in discriminating between /h/ and /ħ/ which will be the task of our subjects in the next chapter. Ghowail (1987) first gathered acoustic data from 5 male NS of Egyptian Arabic in /Cæ/, /æC/ and /CæC/ formats. He first separated the noise portions from the transition and steady state vowel portions and played the vowel only portions to his 10 subjects who came from 4 different Arabic speaking countries. Although his exact methodology is unclear, and specific figures are not reported, he claims that subjects were able to identify the missing consonants based on the vowel portion. Furthermore, when he swapped noise and vowel portions from /hæ/ and /ħæ/ and from /æh/ and /æħ/ tokens, he indicates that the subjects relied on the information provided by the vowel, rather than the noise. When subjects were presented only with a noise portion and a steady state vowel without transition information, they could not identify the consonant accurately. Ghowail's study suggests then that the critical acoustic cue in discriminating between /h/ and /ħ/ is the transition from noise to steady state vowel.

Finally it is worthwhile to consider how the English /h/ is articulated and what acoustic cues may be used in identifying it. The English /h/ is exceptionally difficult to define acoustically as it is offered as an example of extreme underspecification, which, based on the research above appears to also be true of Arabic /h/ and /ħ/. This means that it is heavily influenced by the sounds that surround it and cannot be defined in isolation. Keating (1988) offers the example of /h/ in the phrase "say hoyd again" versus "say hoed again" in which the formant frequencies seen in /h/ interpolate between the surrounding vowels. Similarly Ladefoged (2005) refers to /h/ as a voiceless version of neighboring vowels. He states that it is not really a fricative since the source of noise is not the consequence of being forced through a narrow gap, as for example with /s/ or /ʃ/, but rather turbulence caused by air forced over the edges of the vocal cords and other surfaces of the vocal tract. The noise comes from deep within the vocal tract and thus the resonances of the entire vocal tract are prominent. The sound then is that of a noisy vowel. When it occurs intervocalically, the vocal cords do completely stop vibrating so that F1 and F2 are noisy, but visible, and there is higher frequency noise. Word initially, in the word "high" for example. F3 is clear while F1 and F2 are seen only faintly. Between vowels /h/ simply adds frication noise to a vowel-vowel juncture and between a consonant and vowel it also adds frication noise, but does not affect the transition from consonant to vowel (Olive, 1993)

Based on the evidence that we saw for Arabic above, it seems that the primary difference between /h/ and /ħ/ is that the noisy portion of the latter extends lower, all the way to F1 while that of the former stretched only into F2, according to data from Obrecht (1968). It also seems possible that /ħ/ is of higher intensity than /h/ and may typically be longer, at least in some environments. The research by Heselwood & Tamimi (2011) further suggests that there may be differences in the F1 and F2 of the following vowel in comparing /h/ and /ħ/. They note that these formants are closer together following /ħ/ such that they may be perceptually integrated while those following /h/ are not. It is unfortunate that there is no published research that attempts to vary these cues and determine how they influence the perception of native speakers and this is a clear area of need for future research. The result by Ghowail (1987), suggests that consonants with a primary pharyngeal articulation, like those with a secondary one, are identified primarily by reference to surrounding vowels.

1.5 The Process Approach to Listening Comprehension

During the past decade a group of listening pedagogy researchers and theorists have emerged who advocate a rethink of listening comprehension pedagogy to include more focused instruction on listening subskills. These researchers are united by the following beliefs: that there has been too much focus on the product of listening comprehension and too little focus on the process, that development of bottom-up listening skills has been neglected and that there is great value in deconstructing the listening process and providing training and teaching on individual listening subskills. In this section we take a closer look at this new approach, their criticism of standard practice and the remedies that they propose.

The seeds of this new approach were planted in 1983 in a seminal article by Jack Richards on listening comprehension. This article was the most in-depth consideration yet of just how complex the listening process is and the challenges that it presents the learner. Richards first identifies 3 dimensions of teaching listening comprehension – approach, design and procedure. Approach is a reference to the fact that in order to teach listening comprehension one must have a theory about what listening comprehension is and how people learn to do it in a second language. Design starts from this basis and is an "operalization of information and theory into a form that will enable objectives to be formulated and learning experiences planned" (227). This involves assessing the needs of the learner, isolating the micro-skills of the desired behaviors, developing diagnostic tests and forming instructional objectives. This is the most important section for our purposes for it is here that Richards specifically delineates a number of the micro-skills (or subskills) that underlie listening comprehension. He identifies 33 micro-skills underlying conversational listening including:

> Ability to discriminate among the distinctive sounds of the target language . . . ability to recognize reduced forms of words . . . ability to distinguish word boundaries. . ability to guess the meaning of words from the contexts in which they occur . . ability to reconstruct or infer situations, goals participants, procedures . . ability to predict outcomes from events described (228-299)

In procedure we must decide how to manipulate the input a learner receives and the tasks that she performs in order to develop particular microskills. The importance of the article by Richards is that it provides an explicit framework for the teaching of listening based on a more informed understanding of what the listening process involves. This framework remains a valuable and reliable one today for conceptualizing the process of teaching L2 listening and for evaluating techniques for improving it.

Another important contribution to this new approach was made by Sheerin (1987) who claimed:

> If we are claiming to teach listening comprehension, then it is imperative that we provide more help for learners than merely telling them that their answers are wrong and the right answers are X, Y, and Z (129)

Sheerin claimed that although many language listening materials claimed to teach listening, in fact most of them simply tested listening by providing a listening text and comprehension questions to go along with it. Little account was taken of what type of listening was involved and how questions about it should be varied accordingly. Some materials expected verbatim recall of passages, likely to be difficult for even native listeners. There was also little thought given to what type of information or skills a student would need to successfully in interpret the text.

The combined insights of Richards (1983) and Sheerin (1987) paved the way for a new approach to teaching listening involving a focus on the listening process, rather than product, and ways to improve the subskills that underlie it. We will adopt the terminology provided by John Field (2008) who refers to it as a process approach and contrasts it with the previously described comprehension approach. Although all of the research below is united by its focus on listening processes, not all offer the same solutions.

Mendelsohn (1994) is among the first to offer some specific advice for how to more explicitly teach listening. He offers a strategy-based approach that "makes maximum use of a top-down approach where feasible, but also uses bottom-up strategies where they are deemed useful for learners" (15). He makes a number of recommendations for listening comprehension courses. He notes that they should include lots of listening, that the materials should be in authentic spoken English and that they should feature variety in types of listening. It is also important to carefully control the difficulty of the text so that it is challenging, but comprehensible. Pre-listening is also important and should include a discussion of the topic so that listeners' knowledge is activated. They should also be made aware of what they will be listening for and what level of detail they will be expected to recall. Mendelsohn then delineates some specific strategies for how learners can determine setting, focus on the main meaning, form hypotheses and make inferences. In particular Mendelsohn notes the importance of matching listening type to the tasks that we ask listeners to do. It is not sensible to ask learners to recall every detail of a flight announcement since in real life they would likely listen only for details about a specific flight.

Writing several years after the publication of his 1994 book, Mendelsohn (2001) claims that many of the changes he advocated had occurred or were occurring. He notes that the theoretical literature is showing a much better understanding of the listening process, a welcome shift to a focus on process and a better understanding of the factors that affect listening. However, he claims that there is still a large gap between theory and practice.

Goh (2000; 2002a; 2008) is a proponent of the teaching of metacognitive strategies to learners. She advocates a framework that "emphasizes the constructive nature of learning and the important role that L2 learners play in the process of learning to listen" (2008:195). It is concerned with the path of development of a learner and the importance of automatizing the processing of streams of speech. It also recognizes the benefits of collaborative work among peers in learning how to learn.

Goh (2002b) provides a sample lesson outline for a listening lesson, which involves developing specific skills and strategies. The main part of the lesson plan is little different than more conventional plans, but does include more explicit focus on the development on bottom-up skis. Goh asserts that perceptual level tasks are important but should only be done after the main listening task. She suggests identifying a key phonological feature to highlight and writing the text fragment on the board, then playing the fragment over a few times while pointing out the feature. She also suggests giving learners a transcript of the entire listening and encouraging them to notice how words are pronounced and to think about why they do not recognize familiar words.

The most well-developed process approach to teaching is provided by John Field (1998; 2000; 2002; 2003; 2004, but especially 2008). Field advocates a more explicit focus on training learners to perceive speech more accurately (2003) and notes that when listeners are heavily reliant on top-down processes simple misperceptions, such as perceiving "I won't go to London" as "I want to go to London" can cascade through interpretation of a spoken message (2003).

Field (2004) conducted an experiment to determine what language learners would do when perceptual and contextual information were contradictory. He was able to find evidence that in some circumstances learners are willing to allow contextual information to override perceptual information. His subjects were 38 students at a British school for

English as a foreign language that were either high beginners or lower intermediates from a variety of L1 backgrounds. He compared their ability to identify a word at the end of a sentence when the word was highly predictable versus when it was a minimal pair of the highly predictable word and found that in many cases the subjects allowed the contextual information to override what they heard. A more striking result was when listeners were asked to provide a missing word that was expected to be new to them. In one third of these cases the target word was replaced by a phonologically similar, and more common word, regardless of how appropriate that word was to context. In other words, Field claims, the subjects lack of precision, or lack of confidence, in their perceptual skills made it very difficult to identify new words in spoken speech and they often assumed it was a word they already knew, even if the word was a poor match in context.

In his book *Listening in the Language Classroom*, Field (2008) lays out an approach to teaching listening comprehension with a strong focus on the process of listening, the use of classroom listening tasks as diagnostic tools, and the development of bottom-up skills , especially in the early stages of language learning. Field makes note of the traditional distinction between bottom-up and top-down listening skills favoring instead the terms 'decoding' and 'meaning building'. In decoding, information is transformed from one form to another – from acoustic cues to words and sentences. These sentences (for example 'What a pity!') are meaningless without reference to context and knowledge and these are brought to bear in the process of meaning building. The decoding processes noted by Field include:

> Identifying consonants and vowels . . . recognizing syllable structure . . . matching sequences of sounds to words . . . anticipating syntactic patterns . . . using intonation to support syntax" (2008:115).

While meaning building involves:

analogy with similar listening encounters . . . checking understanding . . . making inferences . . . dealing with pronouns . . . selecting relevant information . . . connecting ideas(2008:117).

Field advocates a skill training approach to teaching learners to listen. One first identifies a final goal that a learner wants to achieve (e.g. expert listening) and then attempts to divide the desired behavior into a number of subskills and to practice each of those skills individually before the learner goes on to practice them together. He claims that this allows learners to convert their declarative knowledge (information about the language) into procedural knowledge (the ability to perform a sequence of actions without conscious attention to setting up the sequence). This allows teachers to begin teaching language rather than just practicing it over and over again. A subskills view also allows the teacher to plan a curriculum that allows for development of learners over time. Field notes that there is insufficient research about whether or not a subskills approach can work for listening comprehension.

Since it is directly relevant to the current project it is worth considering what Field (2008) has to say about phonemic perception and the teaching of listening comprehension. In regard to minimal pair training Field writes:

> Clearly perceiving the differences between phonemes is a useful first step towards producing them. But does it bring any long-term benefits in terms of listening development? Does class time allocated to discriminating between phonemes constitute time well spent, so far as the listening teacher is concerned? We should not overstate the usefulness of ear training of this kind, but it certainly has a part to play in a process approach to listening (168)

Finally he notes that:

> The minimal pairs tradition provides a staring point, but we need to extend it to take account of recent insights into how learners internalize phoneme values. (168).

Field goes on to suggest a number of classroom activities that can be used to support the development of bottom-up subskills. For phoneme discrimination he advocates contrast response ("Put your right hand up when you hear a word with /p/ and your left hand up when you hear one with /b/"), dictations, minimal pair tests and unknown word repetition. He also has suggests activities that help learners perceive syllable structure, word boundaries, prefixes and suffixes and lexical access, and for online sentence parsing, key word recognition, and chunking. Field is not concerned with decoding only, but also advocates development of meaning building skills with using familiar activities such as activating schemata, prediction, inferencing and anticipating. Field clearly sees a role for activities that require students to integrate all of the skills at once by engaging in activities like those required of the comprehension approach. His point of difference is that this should not be the only exposure that students have to L2 listening.

1.5.1 Techniques in developing bottom-up processing

In this section we study some other reports on explicit efforts to improve perceptual skills in L2 learners and the benefits that their practitioners claim for them.

This new focus on the perceptual processes that underlie listening has led to a call by some for a return to an age-old listening teaching technique – the dictation. Stansfield (1985) provides a history of the use of dictation as a teaching methodology. He claims that the classroom practice is related to teaching techniques dating back to the Middle Ages in which a master would dictate knowledge to his pupils. He notes that it is strongly associated with the grammar translation method and has been unfairly tarnished by that association. He claims that in fact the results of dictation tests correlate strongly with overall listening comprehension scores. Ducroquet (1979) calls for a return to a use of dictation in the classroom, but in a modified form. She advocates dictations based upon authentic language materials, the introduction of phonetic transcription, variety in the types of materials, and most importantly greater thought to the specific problems that students face when selecting materials for dictation. Morris (1983) refers to dictation as "a much underestimated teaching technique" (121). She analyzes the pattern of errors revealed by a dictation test given to learners at different levels of an EFL class. She claims that the errors reveal that dictation is not just a recording exercise, but that in fact learners are actively creating meaning. They are thus an excellent diagnostic tool for identifying listening problems. She also claims that dictation is a good way to help learners develop their processing speed. Furthermore, the dictations reveal the need for some students to rely more on compensatory top-down mechanisms as made errors that were wholly illogical in the context they were presented.

Kiany & Shiramiry (2009) report results of an experiment with 60 NS of Persian aged 20-25 that were all enrolled in an English language institute in Tehran. At the start of the experiment all had been enrolled for three terms (approximately 100 hours of classroom instruction). All 60 took a listening comprehension test at the start of the 4th term and an identical test at the end. Half of the students were given 11 additional dictation exercises of about 10-15 minutes each over the course of the term and the other half had no additional instruction. They report that the dictation group showed an average improvement of 14% between pretest to posttest while the control group improved by only 6% and that the difference between the groups was statistically significant. Kiany & Shiramiry suggest several possible benefits of dictation that led to this result. It may be that the dictation group was "forced to listen more attentively to decode the speech" (61). This is similar to the earlier suggestion by Tyler (2001) that forcing students to rely on bottom-up processing skills can have long run benefits. They suggest that the act of repeated dictation may help learners improve their memory. This seems less likely, but it may help them automatize their processing, which amounts to the same thing. They claim that it also exposes them to more native speech rhythms, which they are less likely to experience in the classroom. Finally they note that dictation helps listeners diagnose their own problems with listening comprehension. Chun (2010) also reports on a long term dictation experiment in which university level student in an EFL program in Malaysia engaged in fill-in-the-blanks and dictation of authentic texts over the course of a 12 week course. Subjects showed significant improvements in a TOEFL test given before and after the course. Chun also makes note of numerous positive comments about the value of dictation for the learners; however, there is no control group for this experiment so it is unclear how much of their improvement relates to the dictation practice and how much from listening itself.

Another approach to improving learners' perceptual subskills is what Wilson (2003) refers to as 'discovery listening'. He describes a sample lesson plan in which students first listen on their own and then form groups in an effort to reconstruct the text. In comparing this text to the original (in written form), they are asked to note where they may have gone wrong and try to decide why they failed to understand. An important aspect of the reconstruction phase is that it allows the teacher to more directly observe the ways that students may have misperceived the text. As they negotiate with one another it becomes apparent where common difficulties lie and which are more specific to one or two learners. After this learners are given a copy of the text and asked to indicate where they failed to accurately perceive the message and to indicate why according to a teacher provided rubric (e.g. I couldn't hear which sound it was, I couldn't separate the words, I heard the word, but couldn't remember the meaning quickly enough, the word was new to me, I heard and understood the words, but didn't understand the sentence). Wilson stresses the importance of text selection. They should not be so difficult that listeners over rely on top-down strategies, nor so easy that they are not challenging.

Hulstijn (2003) describes a computer program (called 123LISTEN) devised by a colleague that is intended to improve learners ability to segment speech. To prepare the program the teacher selects a text of several minutes. The text should be comprised mainly of words expected to be familiar to the students. The instructor then divides the text into fragments and provides transcriptions of each fragment. When the student listens she has three ways of listening: without text, with delayed text and with simultaneous text. The main value of the program, according to Hulstijn, is to be found in the delayed text mode. The student is encouraged to listen to each fragment as often as desired, try to reconstruct what is being said, revealing the text to check their comprehension, and then listening again to be sure all words are understood. An immediate concern that one might have with such a system is its labor intensive nature for the teacher; however, a very important aspect it introduces is that it removes the process of developing perceptual skills from the classroom and allows each learner to work at her own pace and according to her own needs.

Another example of computer assisted perceptual training is offered by Hoeflaak (2004). He argues that a certain basic level of perceptual skill is required before top-down processes can be engaged by the learners. Hoeflaak's learners are NS of Dutch studying French who frequently have difficulty segmenting the stream of speech. The program presents the learners with listening texts and with fill in the blank exercises in which learners provide words or phrases according to what they hear. The designers of the listening exercise must try to predict possible wrong answers so that explicit feedback can be provided. Hoeflaak notes an important aspect of the program is that it can save time for teachers who can focus attention in class on the development of productive skills. It also is adaptive to student needs since the types of items a student receives can be tailored to earlier diagnostic testing. Weaknesses include the fact that it requires detailed instruction to work the program and that the program has been adapted to a task for which it is not wholly suited (it was originally designed for grammar instruction). It is also very time consuming to maintain and upgrade the courseware and impossible to predict all possible learner errors.

Since we began our discussion of this new approach to listening comprehension with an article by Jack Richards, it seems suitable to end with one as well. Richards (2005) notes that listening is now a core aspect of many language programs and that there is increasing recognition not only that listening is an important skill in its own right, but that its development supports the growth of other language skills as well. Richards also notes a very important distinction between two different goals for listening in a classroom. In some cases the only goal may be comprehension, in which case traditional lesson plans such as the one described by Field (2008), are appropriate. However, if we also wish the students to acquire language, more explicit interventions that call upon learners to use their bottom-up processing skills are warranted. He suggests that listening texts should first be exploited for the purpose of comprehension, and then be used to foster acquisition. He proposes that after a standard comprehension lesson teachers should engage in noticing tasks and then restructuring tasks. Noticing tasks can include re-listening while reading a transcript of the text, doing a fill-in-the-blanks version of the text, or checking off a list of expressions that occur in the text. Having been made aware of some salient features of the text that the teacher has targeted for acquisition they are then given restructuring activities that prompt them to use the new material communicatively. This could involve dialogue practice incorporating items from the text or role plays in which students are required to use new language items from the text. The concern for such an approach that he voices is that teachers may not understand what the purpose of the second (acquisition) phase is, or that the restructuring activities will not be sufficiently engaging for the students.

1.5.2 Criticism of a focus on bottom-up skills

It should be noted that not all theorists and practitioners are convinced that there is value on the bottom-up approach in general, or in spending time improving phonemic discrimination skills in particular. Some criticisms of this approach, and our response to those criticisms, are presented here.

> In a communicative, interactive context, you don't want to dwell too heavily on the bottom-up, for to do so may hamper the development of a learner's all-important automaticity in processing speech. (Brown, 1994:264).

As noted in our previous discussion, a number of theorists have voiced this concern. They recognize the importance of perceptual skills but worry that focusing student's attention on individual words and sounds will encourage them to use bad habits such as online translation. It is a fair point and undoubtedly must be considered. Classroom time is most profitably used training students to use top-down strategies like schemata activation, prediction and inferencing, and also for focusing attention of functional, communicative language that students can profitably learn. Any rote task, such as a minimal pair test, or phonetic training, is best moved outside of class when the presence of a teacher is not necessary to provide feedback. The phonetic training that we propose makes no demands on class time.

> Practice is the most important thing. The more listening the better, and the subskills will take care of themselves as they become automatized (Ridgway, 2000: 184)

As we shall see in Chapter 3 on second language speech perception, it is not necessarily the case that leaners will just figure it out, indeed the assumption that they will may represent an instance of selection bias – those learners that fail to figure it out fail to succeed and are no longer part of the subject population. It is probably fair to say that some learners will be able to learn to perceive even very difficult foreign language phonemes through naturalistic exposure to the language and some fewer may even do so quite quickly, but others will continue to struggle into higher levels of proficiency and this likely represents a drag on their forward progress. The phonetic training program that we propose will be targeted at those specific learners that have tried and failed to learn important phonemic distinctions in Arabic through normal classroom participation. As noted above by Goh (2008), the process approach can be particularly helpful for less skilled listeners.

> *although it may appear to be a good idea to engage students in distinguishing minimal pairs, such ear training activities may not actually be useful for improving communicative listening where the listener is actively engaged in meaning construction. Because words are not heard in isolation but in specific contexts, listeners are often able to use top-down processing strategies, such as inferencing and elaboration, to complete an interpretation even when they do not recognize every word in the input.* (Goh, 2000:69-70)

It is certainly the case that many times a listener will be able to accurately guess what the missing phoneme was in circumstances when they fail to accurately perceive it; however, there are still very good reasons to feel that more accurate phonemic perception is, other things being equal, more likely to lead to successful listening comprehension. For one, if we maintain the assumption that listeners have a fixed pool of attention to devote to a listening task then more accurate phonemic perception places fewer demands on that pool and frees it to perform other tasks. Also, in the event that the target phoneme is part of a word that is new to a listener more accurate perception means that it is more likely that the listener will accurately encode the new word in his lexicon. Furthermore, it sometimes may be the case, particularly for Americans learning Arabic, that the schemata they use to interpret the word will be misleading when applied to Arabic texts.

More robust bottom-up processes will allow them to understand things that are counter to their expectations and help them build new, more appropriate schemata for the culture of their target language. Thus, more accurate perception can plausibly be expected to improve listening as acquisition, rather then solely comprehension. Indeed, given the extent to which language listeners are tuned to perceive the phonemes of their language, as shall be investigated in the nest chapter, it seems hard to imagine that native-like phoneme perception is not of value in the listening process.

Therefore, while recognizing that each of these criticisms are at least partly true, we believe that the phonetic training that we propose avoids the pitfalls described by these criticisms and can be a valuable part of teaching our learners to listen.

1.6 Conclusion

It was the purpose of this introductory chapter to identify and describe a noted problem in Arabic language pedagogy. Specifically the well known difficulty that many native speakers of English have in mastering the phonological system of Arabic resulting in poor perception and inaccurate production of some Arabic phonemes. A pair of Arabic phonemes, the laryngeal /h/ and pharyngeal /ħ/ were identified as the most difficult two phonemes for NS of English to discriminate and nominated as candidates for a phonetic training experiment. The importance of focusing on bottom-up skills such as phonemic identification was established through the review of a new approach to listening comprehension pedagogy that recognizes the importance of these skills.

This process approach also recognizes the need for pedagogy researchers to develop more in-depth knowledge of the cognitive processes that underlie bottom-up listening comprehension. Thus before we can set out to improve phonemic perception in our learners it is necessary to gain a more complete understanding of exactly how it is that listeners convert raw acoustic signals into the abstract cognitive concepts known as phonemes. For this reason, considerable effort will be made to understand speech perception at its most basic level and to also understand why infant children seem to learn to categorize the phonemes of their native language so effortlessly while adult second language learners frequently face frustration and failure, even after long effort, to learn the same thing. What we hope to demonstrate in the next several chapters is that methodologies do exist that can help many learners to improve their perception of Arabic language phonemes in a relatively short time frame.

Chapter 2: Segmental Speech Perception

2.1 INTRODUCTION

In this and the following chapter we are concerned with understanding the task that an adult language learner confronts in learning to associate sounds in the environment with specific L2 phonemes, an integral part of the listening comprehension process. In the previous chapter we discussed the pedagogical merits of improving the ability of our learners to more accurately perceive the individual segments of the language that they are learning. In this chapter we will try to gain a better idea of exactly what a phoneme is and how speakers organize knowledge about phonemes in their minds and use that knowledge to perceive speech. In the first section of this chapter we consider the problem of defining the phoneme, and then outline some of the primary methods that have been used in investigating phonemic perception. In the remainder of the chapter, we will discuss the main findings of this research in an effort to understand how a listener converts speech sounds (and other information) into abstract phonemic representations in the brain. With an understanding of how speech perception works in native speakers, we will then be better able to appreciate the difficulties confronting second language learners, which is the topic of the following chapter.

2.2 THE PHONEME

Since we are concerned with segmental perception, it makes sense to first define what a speech segment, or phoneme, is. As shall become clear, defining the phoneme is not as straightforward as it may at first seem. Furthermore, there is considerable disagreement as to the extent to which phonemes are in fact meaningful units at the cognitive level, as well as disagreement as to the extent that phonological awareness shapes, or is shaped by, learning to read a phonetic alphabet.

It immediately becomes apparent that it is not possible to define a phoneme in terms of specific articulations or acoustic patterns. For example, the way that most NS of English articulate the phoneme /g/ differs depending on the context in which it is uttered. Prior to a front vowel like /i/ the /g/ is pronounced with a velar point of articulation [g] while after a back vowel like /u/ the point of articulation moves further back in the mouth [ɢ] and the acoustic result changes as well. However, we can conclude that these are not two different phonemes because there are no two words in English that differ only in the articulation of /g/. Furthermore, if we conclude that they are separate phonemes this would lead to the somewhat awkward conclusion that the words "goose" and "geese" begin with different phonemes. By convention, when reference is made to the more abstract cognitive representation of a phoneme backslashes are used /g/, but when referring more explicitly to articulations and their acoustic consequences, brackets [g] are used.

Dresher (2011) provides a brief history of attempts to define the phoneme and states that "the basic concept is that of the unity of sounds that are objectively different but in some sense functionally the same." (242). He notes that while some have rejected the existence of the phoneme as a psychological reality, there are several lines of evidence suggesting that it is a meaningful unit of analysis and that phonemes do have a cognitive existence. For one, unintentional speech errors suggest that some level of phonemic analysis exists. Fromkin (1971) notes a number of examples such as "the nipper is zarrow" when "the zipper is narrow" was intended. Furthermore, many cultures have language games that require phoneme level substitutions in words that would not be possible without phonemic awareness (Bagemihl, 1995 in Dresher, 2011). The final line of evidence cited by Dresher is the alternations in phonological patterning that can cause syllables inside a word to be broken apart. He gives the example from Idsardi (2010) in which the name "Ivan", syllabified as i.van, becomes ki.va.nu when following the preposition "k" and notes that "a representation in which syllables are primitives would have difficulty showing how these words are related." (261). Dresher states that in many of the more recent handbooks on phonology there are no separate chapters for the phoneme, but concludes that it remains a meaningful unit of analysis in distinctive feature theory, underspecification theory, and markedness theory, among others.

It is also sometimes claimed that phonological awareness, though real, is itself an artifact of literacy, which enforces a phonological mindset by virtue of the sound letter correspondence. An experiment comparing illiterate and recently literate Portuguese adults (Morais, Cary, Alegria & Bertelson, 1979) found that the illiterate subjects were much worse than the literate subjects in performing deletion (they are given "cat" and their task is to say "at") and addition (they are given "cat" and told to replace the initial /k/ with /b/) tasks. They conclude "the ability to deal explicitly with the phonetic units of speech is not acquired spontaneously". (330). Similarly, it was found that NS of Chinese that were literate only in the non-alphabetic Chinese characters also had difficulty with addition and deletion tasks compared to similar subjects that had learned Pinyin, a system for writing Chinese characters with the Latin alphabet. In reviewing the evidence for the psychological reality of the phoneme, Port (2007) claims "there is no clear evidence for a universal phonetic inventory of letter-like sound types." (351).

Whatever the case may be, even those that have argued against the cognitive existence of the phoneme have acknowledged that at the very least it has been a convenient fiction for conceptualizing language processes and guiding empirical research. Furthermore, since this particular research is focused on Arabic learning outcomes in university students, we may safely assume that all of our subjects have developed phonological awareness, and that, as discussed in chapter one, improvements in their segmental perception and production are likely, other things being equal, to improve their ability to communicate in a foreign language. We next briefly review the empirical methodologies most commonly used in the research on segmental speech perception and then we discuss what such empirical research has discovered about the perception of speech segments

2.3 EXPERIMENTAL METHODOLOGIES FOR EXAMINING SEGMENTAL SPEECH PERCEPTION

Before examining empirical research that attempts to better understand the way that acoustic signals are mapped to phonemic representations in the brain, it is useful to describe the primary experimental methodologies that have been carried out in empirical studies of speech perception. Since the earliest speech perception experiments conducted by the Haskins Lab in the 1950s (described in more detail below) there have been two main task types that have been used to explore speech perceptual phenomena and these same tasks have also been used in perceptual training studies, which will be discussed in Chapter 4. Subjects have usually been asked either to perform identification tasks or discrimination tasks. In an identification task, subjects are presented with a single stimulus and asked to label that stimulus, usually by selecting from a limited number of options. For example, Liberman, Delattre, Cooper & Gerstman (1954) presented a series of synthetic speech sounds varying in their acoustic characteristics and asked subjects to label the sounds as either /b/, /d/ or /g/. Typically, as in this case, subjects must choose at least one of the offered alternatives. In some cases, subjects are asked to rate the category goodness of the stimulus (e.g. Guion, Flege, Akahane-Yamada & Pruitt, 2000).

In a discrimination task, subjects are presented with multiple stimuli and asked to make some judgment about their similarity to one another. In the simplest version, an AX discrimination task, subjects hear two stimuli and must indicate if they are the same sound, or two different sounds. Such tests are frequently used with continua of synthetic speech sounds. Repp (1981) does this with a continuum of sounds ranging from /s/ to /ʃ/. He presented his subjects with two stimuli separated by 500ms and asked them to indicate if the two sounds were the same or different. A more frequently used task is the ABX task in which the subject hears three sounds in a row and must decide if the final stimulus is the same as the first or the second stimulus (e.g. Liberman et al, 1957; Cross & Lane, 1962). Since this can bias the result in favor of more B responses, more recent experiments have used a AXB design in which the subject must decide if the middle stimulus is like A or like B (e.g. Polka, 1995; Højen & Flege, 2006). Ingram & Park (1998) use an oddball paradigm in which the subject must identify which of three sequentially presented sounds is different than the other two. They argue that this is more economical in terms of the required number of stimulus presentations, but places more demands of phonological memory than AXB and may underestimate the ability of subjects to perceive small differences between stimuli. In an identity discrimination task the acoustic stimuli on "same" trials are literally identical. Thus, if a subject can detect any acoustic difference between two stimuli, they should be successful at the task. In a categorial discrimination task the "same" tokens are not identical, but are different examples of the same speech category. For example (Guion et al, 2000) used a modified version of the AXB task in which the subjects heard three syllables, each provided by a different NS, and had to decide which of the three, if any, did not belong in the group. In this situation, the subjects can only perform the task if they can detect phonologically relevant variation between stimuli that vary in phonologically irrelevant ways as well.

In addition to the order of stimulus presentation, it is important to consider the interstimulus interval (ISI) used in a discrimination task. This is the time that lapses between presentations of the stimuli that the subject must compare. Werker & Tees (1984a) compared NS of English performance on two different L2 contrasts in an AX task with a 500ms ISI and the same task with a 1500ms ISI and found that performance was significantly better for the shorter ISI task. They argue that subjects may have access to a memory trace after 500ms, but are forced to rely on higher level phonological encoding when the ISI is 1500ms.

In general, the benefit of identification tasks is that they more closely approximate real world language tasks. However, the choice of labels can be tricky and it is not always straightforward as to which to choose. Discrimination tasks are much simpler in that subjects do not need to learn to associate sounds with symbols as they do in an identification task. Furthermore, discrimination tasks allow experimenters to examine the internal structure of categories and to probe the extent to which different variants of a given phoneme are perceived as different. It is frequently the case that identification and discrimination tasks are used together to gain the most complete picture of subject perception.

Since these experimental methodologies rely on the ability of subjects to follow directions and press buttons, they are not suitable for use with infants and younger children. Nevertheless, methodologies have been developed to investigate segmental perception in the youngest possible subjects. Polka, Juscyzk & Rvachew (1995) and Jusczyk (1997) provide complete reviews of the relative strengths and weaknesses of the differing methodologies. Since we are primarily concerned with adult speech perception, these methodologies will only be briefly discussed below in the section on the development of speech perception.

2.4 IMPORTANT PHENOMENA IN SEGMENTAL SPEECH PERCEPTION

Understanding the task of an L2 learner in acquiring the ability to perceive speech sounds like a native speaker requires first a consideration of the principles that govern speech perception in a first language. In what remains, we will discuss each of the following – that speech perception is categorical; that listeners rely on multiple, and not necessarily acoustic, cues in speech perception; that incoming speech must be normalized for accurate perception of segments; that the mapping between sound and phoneme is context dependent, and that while human infants are language general perceivers, human adults are language specific perceivers.

2.4.1 Categorical perception

An important concept for understanding the task of the second language learner in acquiring the ability to perceive new segmental speech sounds is that of categorical perception (CP). According to one researcher "Categorical perception occurs when the continuous, variable, and confusable stimulation that reaches the sense organs is sorted out by the mind into discrete distinct categories" (Harnad, 1987; ix). This is best exemplified in the nonspeech realm by color perception. The refracted light of a rainbow exists along a spectrum of different wavelengths, and if our perception of color were continuous, we would perceive a color blend from red to violet, rather than distinct bands of color. What this means is that the physical difference between two wavelengths and the perceived difference between them by a human observer, is not correlated. Wavelengths that are assigned by an observer to the same category (e.g. "red") should be more difficult to tell apart than two wavelengths assigned to two different categories ("red" and "orange) even if the physical difference between the former pair is greater. A similar process occurs in speech perception, but it is not so easily observed. In this section we will review important findings about the categorical perception of speech sounds for native speakers, mainly of English, in preparation for a discussion in the next chapter about the task facing a second language learner in acquiring a new speech system.

The discovery of categorical speech perception was among the earliest findings of the Haskins Laboratory, a famed center for speech perception research. Their earliest research was made possible by the invention of the pattern playback device (Cooper, Liberman & Borst, 1951), which allowed them to draw spectrographs that they could then convert into sound. They quickly found that although spectrographs of actual speech were finely detailed, much of this detail could be left out when constructing synthetic speech sounds. Although these more idealized representations did not sound like natural speech, they were sufficient to reliably cue the perception of English language phonemes (see Liberman, 1996 for a history of the Haskins lab and its main findings).

In the first ever test for CP of speech sounds, Liberman (1957) created a continuum of sounds with variation in the second formant transition that his lab had previously determined to cue the perception of the voiced stops /b/, /d/ and /g/. Thus the stimuli existed on a 14 step continuum from /b/ through /d/ to /g/. They had their subjects perform two tasks. In the first they performed an ABX discrimination task in which they are presented with stimuli from different points along the continuum to see if perception is more sensitive in some areas than others. In the second they heard a series of the synthetic sounds and were asked to label them as /b/, /d/ or /g/. Idealized representations of the pattern of results for a group of subjects are shown below. The figures use idealized data, but are based on Liberman et al (1957).

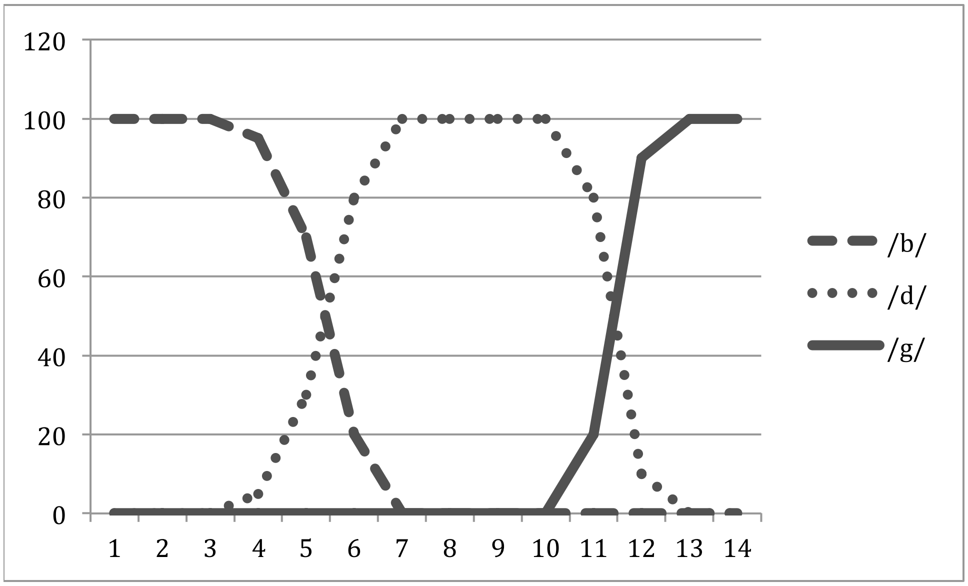


Figure 2.1: Idealized Identification Data

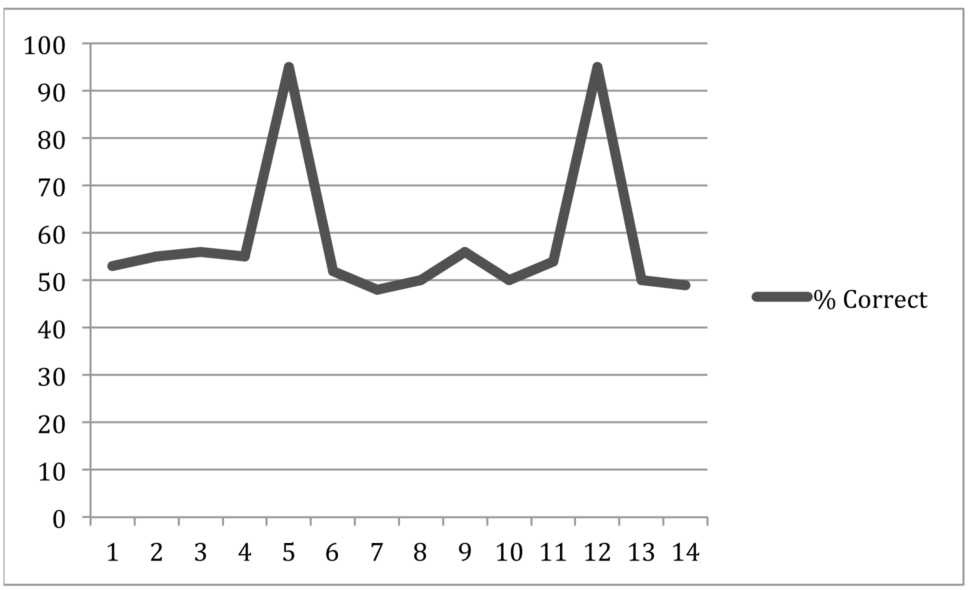


Figure 2.2: Idealized Two-Step Discrimination Data

In Figure 1, note that the first 4 stimuli are consistently identified as /b/ with a sharp drop off immediately thereafter such that stimuli 7-14 are never identified as /b/. A similar pattern is shown for the other two phonemes as well with well-defined regions of stimulus space that reliably cue either /d/ or /g/. Were the stimuli perceived continuously, we would expect that the slope of the lines separating the regions would be considerably less shallow. In Figure 2 we see idealized data from a two-step discrimination test. These data show that the subjects were very poor at discriminating between two stimuli that they had previously identified as belonging to the same category, but very good at discriminating between two stimuli that they had previously identified as belonging to two different categories. Note that in the discrimination task, which was given first in Liberman et al (1957), the subjects were not told that the stimuli were to be perceived as speech, and were told to focus on any noticeable difference between the stimuli. In perfectly categorical perception identification results should predict discrimination results. Subjects should fail to discriminate any two stimuli that they place in the same category, but should succeed in discriminating between any two stimuli they place in different categories.

This pattern of results - sharp boundaries in the identification function and peaks in the discrimination function - are the hallmarks of categorical perception and the results of this experiment have largely held up over time. That we perceive at least some speech sounds categorically is an uncontroversial statement. Categorical perception has been demonstrated for a number of different speech continua including VOT, place of articulation, liquids, fricatives and vowels (see Repp, 1984 for a review). Repp (1984) does note that "the perception of liquids, semivowels, and fricatives is clearly less categorical than that of stops". (288)

A recent study (Chang, Rieger, Johnson, Berger, Barbaro & Knight, 2010) recorded cortical local field potentials in the posterior superior temporal gyrus (pSTG) from subjects undergoing craniotomy (a surgical opening of the skull). Subjects were given a classic oddity discrimination task with a /ba/-/da/-/ga/ continuum and then exposed to the same sounds while the cortical readings were taken. The investigators found "that the neural response patterns were strongly organized along phonetic categories and did not demonstrate sensitivity for gradual acoustic variation". (1428) The study demonstrates fairly conclusively that at least at some level the human brain is responding categorically to speech sounds.

The claim by Liberman that categorical perception is unique to speech, and the claim that it is evidence for a special neural module that matches speech sounds to the motor patterns used to generate them, has not been accepted uncritically and has been the subject of substantial research. Among the important questions of that research were: How categorical is the perception of speech sounds? Is CP unique to the perception of speech sounds only? Is the CP of speech sounds unique to humans? What crosslinguistic similarities and differences exist in the categories of speech sounds? Can the pattern of results be altered through experience with speech sounds?

The first of these questions, about the extent of CP of speech sounds, relates to the finding that although intracategory discrimination was poor, it was not wholly absent. For example, Pisoni (1973) found that by shortening the interstimulus interval (ISI) in a discrimination task, subjects perception of vowel stimuli was significantly less categorical, he suggested that this is due to a second, acoustic mode of speech perception which allowed subjects to briefly retain the acoustic information in their memory and thereby compare it exactly with a new stimulus. However, under longer ISIs, subjects could only retain the phonemic label that they had assigned to the sound, and perception was then categorical. Additionally, Pisoni & Tash (1974) found that reaction times in comparing two stimuli that were acoustically different, but given the same phonemic label, were longer than those when comparing two acoustically identical sounds, suggesting again that perception of these sounds was not wholly categorical. It has also been found that how subjects are instructed and the specific discrimination task that they perform can alter the degree to which perceptual tasks show evidence of categorical perception (Pisoni & Lazarus, 1974). It has also been found that training with feedback can enhance the ability of listeners to perceive acoustic differences between sounds that are given the same label (Carney, Earley, Widin & Viemeister, 1977), a point to which we will return in Chapter 4.

The internal structure of phonetic categories has also been explored with the use of category goodness judgments. For example, they take a typical bilabial VOT continuum and extend the series well beyond a typical /p/ sound and ask subjects to rate on a scale of 1-10 how good of a /p/ each member of the continuum is (Wayland & Miller, 1992 cited in Miller, 1994). The data show that only a narrow range of VOT values receive the highest ratings. Further studies have shown that category centers, like category boundaries, shift according to speaking rate (Miller & Volaitis, 1989) and place of articulation (Volaitis & Miller, 1992). These experiments further demonstrate that phonetic perception is not fully categorical since subjects do differentiate between members of the same category, at least in some laboratory tests.

Among the original claims of the Haskins group was that CP was special to the perception of speech sounds and that the perception of nonspeech sounds was always continuous. However, experiments showing categorical perception of nonspeech sounds, and experiments showing that some nonhuman animals perceive speech sounds categorically, demonstrate that this is clearly not true. For example, it has been found that subjects demonstrated categorical perception of a continuum of “noise-buzz” sounds, designed to mimic the temporal difference that distinguishes voiced from voiceless consonants (Miller, Wier, Pastore, Kelly & Dooling, 1976). A series of experiments using a continuum from “plucked” to “bowed” noises (Cutting & Rosner, 1974; Rosen &

Howell, 1981; Cutting, 1982) demonstrated that in at least some circumstances these stimuli are perceived categorically and also demonstrated the "general fickleness of the phenomenon of categorical perception (Cutting, 1982: 474). At the very least it does not seem that CP is always unique to the perception of speech sounds. Evidence from training studies, reviewed in detail in chapter 4, suggest that it may be the behavioral relevance of the perceptual task that determines whether or not CP will develop (e.g. Guenther, Nieto-Castanon, Tourville, Ghosh, 2007).

Further evidence against the notion of CP as a unique property of speech perception comes from studies of how other animals perceive human speech. Since their pattern of results cannot possibly be driven by speech processing, any evidence of CP of speech sounds in animals argues against CP as a special feature of speech perception. Kuhl & Miller (1978) trained chinchillas to respond differentially to endpoints of a VOT continuum and then exposed them to the full continuum to test for categorical perception. They found that the identification function was categorical (though shallower than for human subjects) and that the boundary point shifted along with the place of articulation, just as with humans. Kuhl (1981) tested the discrimination ability of chinchillas and found peaks in discrimination at category boundaries, just as with human perpception of these sounds. Taken together, the two experiments reveal that chinchillas perceive these speech sounds in much the same way that humans do. The experimenters propose that human language developed to take advantage of naturally existing discontinuities in perception that are common to all mammals that categorical perception is an artifact of those discontinuities, albeit a very useful artifact.

Cross-linguistic differences in how speech sounds are categorized are of obvious interest to foreign language teachers and learners. If language developed to take advantage of preexisting psychophysical boundaries in acoustic space, we would expect that all languages would use the same boundaries. Lisker & Abramson (1964) examined spectrographic measurements of VOT in word initial stops from speakers of a variety of different languages. There were crosslinguistic differences in the VOT measurements of voiced and voiceless stops, but these differences were not completely random. There were three different zones around which VOTs clustered. For example, in looking at VOT of labial stops they found that values for NS of English clustered around either a VOT of 0 or a positive VOT of around 50. However, for NS of Puerto Rican Spanish they found that VOTs were either negative (with a large range of different negative values) or around 0. For Thai speakers there were peaks at 3 locations, consistent with the existence of three bilabial stops in that language. Subsequent perceptual studies (Lisker & Abramson, 1970 cited in Abramson & Lisker, 1973; Abramson & Lisker, 1973) showed that the perceptual boundaries for perception of voicing were predicted by the production data. Thus, although English and Spanish both have voiced and voiceless bilabial stops they categorize a continuum of synthetic sounds differing in VOT in a different manner.

The final question regarding CP, as to whether or not experience can alter how speech sounds are categorized, will be considered in each of the next two chapters on second language speech learning and phonetic training.

2.4.2 Perceptual cues

Speakers of a language do not rely on a single cue in mapping from sound to phoneme but must integrate information from multiple sources. Lisker (1986) lists 16 different “pattern properties” that may play a role in cuing perception of voicing on a bilabial stop (i.e. /b/ or /p/) in a VCV context including duration of closure, duration of preceding vowel and intensity of the release burst. Bohn & Flege (1990) created a series of synthetic stimuli ranging from “beat” to “bit” and another from “bet” to “bat” that varied in both the quality and duration of the vowel. They presented these 33 stimuli to NS of English and asked them to identify each stimulus as an example of “beat” or “bit” (or “bet or “bat”). They found that NS of English relied primarily of the quality of the sound in determining which vowel sound they were hearing, but also on the duration of the vowel. Other things being equal, a longer vowel was more likely to be labeled /i/ than /ɪ/.

As noted in the Lisker study above, acoustic cues that might generally be considered to be part of another phoneme are used by listeners to detect the voicing feature. Raphael (1972) found that voicing was perceived in part by reference to the duration of the preceding vowel with longer vowels more likely to lead to the perception of voicing. Raphael (2008) contains a list of the variety of acoustic cues that have been found to cue the perception of segmental speech sounds according to the place and manner of articulation. It is important to note that although the patterns of contextual variation in the realization of different allophones of a phoneme is not completely random, it does vary from language to language. For example, in Finnish, unlike in English, the duration of vowels preceding consonants is not a reliable cue for voicing (Flege & Hillenbrand, 1986) and even different dialects of the same language may differ to the extent that vowel duration is a reliable cue to vowel identity (Escudero & Boersma, 2004).

Remarkably, the cues relied upon in assigning an input signal to a phonemic category are not exclusively auditory. This is demonstrated by the McGurk Effect (McGurk & MacDonald, 1976). In this experiment the researchers made videos with mismatched acoustic and visual information. Subjects saw a video of an individual saying /ga/ but heard audio of the same person saying /ba/. There was significant variation in the results of the experiment, but most subjects reported hearing /da/, apparently a compromise between the velar /g/ and bilabial /b/. Subsequent experiments have found that the magnitude of the effect is influenced by the vowel following the consonant (Green, Kuhl & Meltzoff, 1988). In summarizing research on the time course of the integration of audio and visual information Rosenblum (2008) concludes that "audiovisual speech is integrated at one of the earliest possible stages that can be observed" which in turn is "consistent with a speech function that is relatively unconcerned with modality." (54) Auditory information is not privileged in speech perception.

2.4.3 Allophonic variation

Another concept critical to understanding the task of the foreign language learner is that of allophonic variation, which was discussed briefly in the introduction to this chapter. Phonemes have different articulatory and phonetic variants according to phonetic context or speech register. An example with the English language phoneme /t/ is given by Hayes (2009:121-124). The articulatory realization of /t/, which is an aspirated alveolar stop in initial position ([tʰ]), frequently (though not obligatorily) becomes preglottalized [ˀt] in the final position. When /t/ occurs intersyllabically, and stress falls on the second of those syllables, it is often realized as a tap [ɾ], as in "butter" and when it is preceded by /s/ it loses its aspiration and becomes [t] - the puff of air following /t/ in "top" is absent in "stop".

From a perceptual standpoint, this presents a challenge for the listener who must be able to undo these allophonic changes and recover the intended phoneme. This is especially true since the rules for allophonic variation can cross word boundaries so that a given word could sound very differently in two different acoustic contexts. It does appear to be the case that there is a cognitive reality to the phoneme, that is that listeners do recognize that each of the phonetic variants of /t/ above belong together as single phoneme. They do this despite the fact that the same acoustic pattern may cue the perception of two different phonemes according to context. For example, Liberman, Delattre & Cooper (1952) found that the same exact noise burst was perceived as a /p/ when followed by /i/ or /u/ but as a /k/ when followed by /a/. The importance of context is also demonstrated by Pegg & Werker (1997). In this experiment subjects were asked to give category goodness judgments for [d] and [t] (the unaspirated version of /t/ found in “stop” but not “top”) and it was found that even though the two acoustic patterns came from different underlying phonemes, subjects perceived no reliable difference between them. In an AXB discrimination task, the subjects showed some ability to detect the difference, but it was far below that for typical interphonemic discrimination. Thus we find that listeners are sensitive to phonetic context in perceiving allophones and that they have difficulty discriminating between allophonic variants of a single phoneme. As shall be seen, this fact has implications for second language speech learning.

For the purpose of the foreign language learner the importance of allophonic variation is that single phonemes have multiple different acoustic realizations, depending on the context in which they are uttered, and that perception of acoustic patterns depends on the phonotactic rules of the native language.

2.4.4 Normalization

In addition to the variability introduced by phonotactic constraints, which are typically shared by all speakers of a given language variety, listeners must also account for acoustic variations that are the result of speaker specific factors such as differences in the size and shape of the vocal tract. Peterson & Blarney (1952) recorded American

English vowels from 76 different NS of English (men, women and children) and plotted the first and second formant values obtained. They found that there was substantial category overlap. That is, judging by these values alone, it was not possible to reliably assign each to a distinct vowel category. However, when second group of subjects was asked to identify the vowels in each of the tokens, their accuracy was much better than would have been expected. Despite the overlap, speakers and listeners agreed about what vowel was spoken. Thus it seemed that perceivers were modifying their judgments according to information they were able to gather about the speaker.

Ladofoged & Broadbent (1957) synthesized several different versions of the sentence "Please say what this word is . . ." in which the vowels in the sentence were varied to have either higher or lower first and second formants. They also synthesized several different /bVt/ tokens in which F1 and F2 were varied. They played the sentences followed by the word tokens and asked their subjects to identify the word that they heard at the end of the sentence. As expected, they found that the identity assigned to a given /bVt/ token varied according to the sentence that preceded it. Thus, the same exact acoustic signal mapped to different phonemes based on prior information about the identity of the speaker and the pattern of subject responses suggested that the alternations were made based on the different F1 values of the carrier sentence. Subsequent studies have confirmed this basic finding.

The exact mechanism by which this normalization process occurs is not precisely understood. Sussman (1986) offers a neuronal mechanism postulating that human vowel identification is similar to bat echolocation, which like speech perception requires the integration of complex acoustic information. A first layer of neurons, each responding to a different frequency, feed forward into a second layer that is combination sensitive, receiving input from earlier layers and calculating the ration of, for example, F1 and F2.

This ratio, rather than the absolute values, are what cue the eventual precept of /i/, /a/ or /u/. This type of model is true normalization in that the stored abstract representation is shorn of any irrelevant detail and the same must be done with incoming stimuli so that they can be compared with stored exemplars. Formant ratio models such as this have been criticized as incapable of explaining the fact that listeners can accurately identify vowels in whispered speech, despite its lack of a fundamental frequency (F0), which should be necessary for establishing the requisite ratios.

Other attempts to explain the phenomenon demonstrated by Ladofoged and Broadbent (1957) argue that the abstract representations stored in listeners' minds include information about every instance of a given phoneme and that phonemes are identified according to how closely they match these richly specified representations (see Johnson & Mullennix, 1997 for a review of this exemplar-based theory). So rather than having a simple stored representation and a complex process whereby that representation is accessed, these researchers propose that the representation is complex in that it somehow reflects the characteristics of all perceived exemplars of a given phoneme, such as speech rate and the characteristics of individual speakers. A more parsimonious approach is that taken by neural network models such as Guenther & Gjaja (1996). In these models, cognitive representations are shaped by exposure, but do not encode details from each instance of a given phoneme.

2.4.5 Infant speech perception

As we are concerned here with speech learning in adults, we will consider the results of speech perception in infants only briefly. The most important finding of infant speech perception for our purposes is that infants up until about the age of one year are language universal perceivers. In other words, infants can reliably discriminate between the phonemes of any language, regardless of the linguistic environment in which they are being raised. This means that when we attempt to train adults to perceive new phonemes we are not necessarily attempting to give them new abilities, but perhaps to uncover old ones.

The most thorough experiment to demonstrate this is provided by Werker & Tees (1984a). They conducted a longitudinal study and tested infants at three different ages – 6- to 8-month olds, 8- to 10-month olds and 10- to 12-month-olds on perception of the Hindi /ta/-/ʈa/ contrasts and the Nthlakampx* voiceless glottalized velar versus uvular contrast /k̰i/-/q̰i/. At 6-8 months all 6 infants were able to discriminate both contrasts, at 8-10 months all of them were able to make the Hindi contrast, but only 3 could for the Nthlakampx contrast and by 10-12 months none showed evidence of discrimination for either of the contrasts. Based on these data, Werker & Tees conclude the perceptual tuning to the native language occurs during the second half of the first year of life. An additional follow up study (Werker & Lalonde, 1988) used synthetic stimuli along a /ba/-/da/-/ta/ continuum and found that 6- to 8-month-old infants discriminated these stimuli categorically, that is they responded to changes that were phonemic in either English or Hindi, but not to those that crossed no phoneme boundaries. Once again, infants in the 11- to 13-month old group were only able to discriminate stimuli that crossed a boundary that was phonemic in their own language (English). Similar experiments have shown that Japanese learning infants between 6-8 months, but not 10-12 months, can discriminate the /r/-/l/ contrast of English (Tsushima, Takizawa, Sasaki, Shiraki, Nishi, Kono, Mneyuk, & Best, 1994) and that Kikuyu learning infants (1-4 months) can discriminate the /b/-/p/ contrast, which is not phonemic in their language (Streeter, 1976).

* the language is referred to as Thompson in the original paper, but subsequent references refer to it as Nthlakampx

Although we can be confident in stating that infants become phonetically tuned to their native language during the first year of life, it is considerably more difficult to determine exactly what drives this process. Nevertheless, there appears to be some broad agreement in the field about what is happening. The process can be divided into three overlapping phases. In the first, newborn infants are innately endowed with predispositions that direct their attention to speech sounds and certain aspects of those sounds. Next, exposure to these sounds results in the formation of phonetic categories based on the distribution of sounds in the input. Finally the onset of word learning further refines those categories and specific sound to phoneme mappings are learned.

2.5 Conclusion

This chapter has considered the phenomenon of speech perception as it has been explored in native speakers. As we have seen, for a listener to accurately reconstruct an intended message requires a number of varied skills that must be engaged simultaneously and rapidly. The listener must be able to match incoming acoustic and visual signals with stored and very abstract representations in the brain. They must do this while confronted with substantial variation of two different kinds. On the one hand, the relationship between sound and phoneme is heavily context dependent such that the same sound may map to two or more different phonemes, depending on context, or two very different sounds may map to the same exact phoneme. At the same time, the listener must cope with the linguistically irrelevant variation in the speech stream introduced by the fact that different speakers have different sized vocal tracts and different ways of producing speech sounds and who speak at different rates.

The importance of this for our purposes is that it is not just the individual phonemes that vary from one language to the next and that present problems for our learners. Languages also vary considerably in how phonetic context influence the realization of phonemes and our learners must learn to decode speech according to these patterns or the likelihood of miscomprehension increases. In the next chapter we consider some of the individual and linguistic factors that influence the process of second language speech learning and describe the models that have been developed to explain and predict the outcome of that process.

Chapter 3: Segmental Speech Perception in an L2

3.1 Introduction

In this chapter we consider the task of an adult L2 learner that needs to learn how to map sounds in the environment to phonemic representations in the brain as part of the process of learning to communicate in a second language. In Chapter 2 we saw that the process of speech perception involves the assignment of auditory patterns to abstract phonemic representations in the face of substantial variation in the input signal. Although there is substantial overlap cross-linguistically in the use of category boundaries and patterns of allophonic variation (see for example Lisker & Abramson, 1964) there are also cross-linguistic differences in how speech sounds are patterned and therefore to comprehend the speech signal an L2 user must learn, among other things, a new pattern of speech categories and their numerous varieties and learn which information in the speech signal to focus on, and which to ignore. In this chapter we review research that helps to make inferences about the specific factors that influence the outcome of segmental speech perception in an L2. Where relevant, evidence from phonetic training studies will be presented; however, a full discussion of phonetic training is reserved for the following chapter.

3.2 Factors affecting L2 speech learning

We can divide the factors affecting the ability to perceive L2 segments into two main types. Context factors relate to the learning situation – at what age does learning begin, to what extent is the L1 still actively used, what is the quality and quantity of L2 input. Language factors are influenced by the relationship between L1 and L2 phonetics and phonology. Since we will assume that all of our learners are in more or less the same context, we will only briefly treat those factors and how they relate to our learners and then discuss language factors in more depth.

3.2.1 Learning context factors

One important factor in predicting the success of L2 segmental production and perception is the age at which exposure to the L2 begins, usually operationalized as the age of arrival (AoA) of the subject in an L2 environment. Flege, Munro & McKay (1995) elicited speech samples from 240 NS of Italian - all of whom were presumed to be at L2 endstate, but varying in their age of arrival in Canada – and had NS of English rate the samples for accentedness. There was substantial intersubject variation, but a clear pattern emerged: the speech samples of earlier arriving subjects were, on average, judged to be less accented than later arriving subjects. Note though that AoA correlated inversely with other potential explanatory factors such as length of residence and amount of L2 use (Flege, 1998 cited in Flege & Liu (2001)). A similar study (Flege, Yeni-Komshian & Liu 1999) was conducted with NS of Korean with varying AoAs to the United States. Once again, the accents of 240 subjects were rated and compared with AoA. This study also carefully controlled for other potential confounding factors such as length of residence, amount of English language education, and amount of L2 use. It was once again found that, on average, the speech of later arriving subjects was rated as more accented. Flege, MacKay & Meador (1999) examined the production of English vowels by NS of Italian that varied in the arrival to Canada and found a similar affect. Subjects in the late arrival group pronounced 6 of 10 vowels worse than NS, those in the mid group were worse on only one, and those in the early group on none.

Studies examining the perception of L2 segments are rarer, but are more or less in line with the production studies – earlier is better. The same study that measured the production of English vowels above (Flege et al, 1999) also tested the ability of the Italian subjects to discriminate English vowels. The early arrival group was indistinguishable from the NS while the mid group performed more poorly on 4 vowels and the late group on all 8 tested vowels. MacKay, Flege, Piske & Schirru (2001) compared late and early Italian learners of English in their ability to identify English voicing contrasts with the stop consonants. They also divided their subjects according to the extent to which they still used their L1 regularly. They found that only the late-arriving, high-L1 use group was significantly worse on the task than NS of English. Yamada (1995) examined discrimination and identification of synthetic versions of English /r/ and /l/ by NS of Japanese varying in their first exposure to English in an immersion environment, and the length of residence in that environment. She concludes that the younger the subjects were, and longer the subjects were exposed to English the better they became at discriminating and identifying English /r/ and /l/. Flege, Takagi & Mann (1996) also compared inexperienced (mean LOR = 2 years) and experienced (mean LOR = 21 years) NS of Japanese in their identification of word initial /r/ and /l/ in naturally produced English minimal pairs and found that the experienced group was significantly better than the inexperienced group, though significantly worse than NS. They also found a lexical familiarity effect. They compared subject performance in identifying /r/ and /l/ when one member of the pair was a word and the other not, with identification of just the isolated /l/ or /r/. They found that subjects were more accurate in identifying real words than in identifying the isolated beginnings of those words, suggesting that the subjects were using their knowledge of English lexicon to bias their decisions. NS of English identified all the stimuli at near 100% indicating they we able to perceive the difference in the sounds regardless of lexical context.

Although early exposure and lengthy exposure tend to go hand in hand, Tees & Werker (1984) found that a group of subjects that were raised in a Hindi environment until about age 2, and then had no further exposure to the language, were able, at around age 25 to perceive the difference between Hindi dental stop /t/ and the unvoiced aspirated retroflex /ʈ/, while NS of English without such experience could not do so, even after a year of Hindi study. They could do this even though they had no ability to speak or understand Hindi, and performed no better than their classmates with no prior exposure to Hindi. Oh, Jun, Knightly & Au (2003) compared 3 groups of college aged Korean learners on their ability to perceive and pronounce a 3-way Korean voicing distinction – a plain /t/, an aspirated /t^h/ and a tense /t'/. The groups were novice learners (NS of English), childhood speakers (they had spoken Korean for at least 3 years in childhood, but then suddenly stopped by age 7) and childhood hearers (they had heard, but not spoken Korean in childhood). Once again, the benefits of early exposure were confirmed, both the childhood speakers and the hearers outperformed the novice learners on a perception task and the childhood speakers, but not hearers, outperformed the novice group in a pronunciation task. However, a similar study (Ventureyra, Pallier & Yoo, 2004) failed to confirm this result. The main subjects in this study were individuals that were adopted from Korean into French families and had no further exposure to Korean. The subjects in Oh et al (2003) had come from the immigrant community and had been enrolled in a Korean class. Ventureya et al tested a similar voicing contrast (but a different place of articulation) and found no advantage for the early exposure group compared to NS of French. They suggest two possible reasons for the discrepancy. It may be that the subjects in the Oh et al study maintained their perceptual abilities because they continued to be exposed to Korean, while the Ventureya et al subjects did not. Alternatively, it may be that the Oh et al subjects, having returned to the classroom, had uncovered their abilities after a few months of exposure and that the same would be true for the Ventureya et al subjects.

Although it seems to clearly be the case that earlier is better, research with Spanish-Catalan bilinguals suggests that even intensive, early exposure to an L2 may be insufficient to yield perception patterns similar to those of monolingual natives. Pallier, Bosch & Sebastián-Gallés (1997), Sebastián-Gallés & Soto-Franco (1999) and Bosch, Costa & Sebastián-Gallés (2000) each show evidence that there are important differences between NS of Catalan and Spanish/Catalan bilinguals, even for very young infants (Bosch et al, 2000).

Four important points stand out regarding the age of exposure to L2. First, it seems clear that other things being equal, the earlier one is exposed to the L2 the more nativelike one's L2 speech perception is likely to be at asymptote. Second, that it cannot be assumed that early, intense exposure to L2 necessarily yields nativelike perception. Third, that early exposure alone, even in the absence of continued exposure to the language may be advantageous to later learning. And finally, although clear trends do emerge from the data, there are notable exceptions to these trends. For example, in the Flege et al (1995) study there were instances of late arrivals with low accentedness ratings and early arrivals with high ratings.

While the primary factor determining the ability to pronounce and perceive second language phonemes seems to be the age at which L2 learning began, a number of other factors have been found to influence the success rate of L2 segmental attainment. A second factor found to be important in predicting L2 perceptual abilities is the ratio of L1 and L2 use. Since the two tend to inversely correlate, it is difficult to determine whether continued L1 use, limited L2 use, or both are responsible, but in general, the more an individual is using her L2, the more poorly she will learn L1 speech contrasts. The Flege et al (1995) study mentioned above found that age of arrival was the most important factor in the accentedness of NS of Italian that had moved to Canada, accounting for 63% of the variance across 240 subjects, but it also found that language use factors accounted for about 15% of the variance. Flege et al (1999), which examined accentedness ratings of NS of Korean speaking English, found that the most important factor was again age of arrival, but when controlling for AoA they found that reported use of L1 and L2 was also a significant factor. In general, the less frequently the subjects used Korean, the less accented their English speech was. Studies measuring the perception of L2 phonemes based on the ratio of language use are rare, but we saw above that MacKay et al (2001) found that NS of Italian that arrived early in life perceived English voicing contrasts accurately, but late arrivals did so only if they reported low use of L1. They also found a learning advantage for subjects with higher scores on tests of phonological short-term memory (PSTM). A study by Flege, Bohn & Jang (1997) compared perception of the English vowel pairs /i/-/ɪ/ and /æ/-/ɛ/ by NS of German, Mandarin, Spanish and Korean and looked for differences in which acoustic cues experienced (mean LOR = 7.3 years) and inexperienced (0.7 years) learners used in identifying the vowels. They found that experienced, but not inexperienced learners used more nativelike strategies (i.e. reliance on spectral rather than temporal information) in identifying the English vowels. Flege & Liu (2001) attempted to account for the quality of the linguistic interactions that language learners have in their L2. They note that prior studies have failed to find an effect for length of residence (LOR), a proxy measure for the quantity of L2 input, which supports the idea that only AoA matters. Since younger learners are much more likely to be educated in the L2, and older workers more likely to take employment where they continue to use the L1, the age differences seen in other studies may be due to the fact that younger learners are getting more and higher quality L2 input, rather than the effects of age on the ability to learn new phonemes. They tested NS of Chinese living in the United States and they divided their 60 subjects into 4 groups. There was a long LOR group (3.9-15.5 years) and a short LOR group (.5-3.8 years) and each of these groups was divided into two groups of students and nonstudents. The subjects' perceptual task was to identify word final stops (/b d g p t k/) in English pseudowords. Chinese has very few stops in word final position, and this task has proven difficult for them in the past. As expected, the researchers found an interaction between LOR and the input condition. For the nonstudents, LOR was irrelevant, those with short and long LOR performed nearly identically. However, for the students the effect of LOR was significant and the long LOR group outperformed the short one. As might be expected, individuals that use their L2 more actively show greater perceptual learning.

3.3 Linguistic factors

The factors reviewed above are likely to be generic to any combination of L1 and L2. The specific difficulties that a language learner will have in attempting to acquire a new speech perceptual system are dependent upon the relationship between the L1 of that learner and the L2 that she is trying to learn. As the experiments reviewed below will demonstrate, learners of a new language, at least on initial contact, will filter the speech sounds of L2 using the well-entrenched perceptual routines of the L1. In what follows, we review empirical evidence that helps to understand why some L2 sounds are easier to learn than others.

3.3.1 Empirical studies revealing difficulties in L2 perception

Difficulties in the acquisition of L2 pronunciation date back at least to the bible (Judges 12:5-6 New Jerusalem Bible) and the Arabic grammarian Sibawayh mentions the difficulty that many non-Arabs have in pronouncing certain Arabic phonemes at least as early as the 7th century. Although evidence of misperception of L2 speech is harder to come by, it has been suggested that some of the changes in letter/sound correspondence that occurred as the alphabet spread to other languages, are the result of misperceptions by those adapting the alphabet from a foreign language to their native one (Huehnergard, 2010). Furthermore, there is substantial anecdotal evidence that L2 learners occasionally struggle to perceive some foreign speech sounds. In this section though, we will focus on empirical studies of perception in the hopes of gaining a more detailed picture of the problems that arise in perceiving L2 speech and why. Goto (1971) and Miyawaki et al (1975) compared Japanese and English speakers ability to identify and discriminate English /r/ and /l/ and found that many of the Japanese were quite poor, while the task was easily performed for the NS of English. Similarly, Werker et al (1981) compared NS of English and NS of Hindi in the ability to discriminate a Hindi retroflex/dental place of articulation contrast and found significant differences between the two groups. The NS of English were very poor at the discrimination. Furthermore, inexperienced Mandarin Chinese and Taiwanese listeners have difficulty discriminating between voiced and unvoiced stops in word final position, where a voicing contrast does not exist in their native language (Flege, 1993). NS of English have difficult perceiving the German vowel contrast /y/-/u/ (Polka, 1995) and Flege (1995a) cites research showing that the size of native speaker vowel inventory influences the categorization of English vowels for L2 users. Other examples include the difficulty that NS of Canadian French have in perceiving the /θ/-/ð/ (Jamieson & Morosan, 1986) and the /d/-/ð/ contrast (Polka, Colantonio & Sundara, 2001) of English.

3.3.2 Differential success in the perception of L2 segments

That adult second language learners sometimes have difficulty perceiving L2 speech sounds will of course come as no surprise to anyone that has taught or learned a second language. Why they have so much difficulty is a harder question to answer and requires more in-depth experimentation. It was initially proposed that the failure of language learners to perceive phonemically relevant acoustic details in a given L2 was due to a loss of sensory ability caused by a lack of use (Eimas, 1975). However, it has been found that under certain conditions adult listeners can in fact detect the acoustic differences between L2 phonemes that they fail to discriminate in minimal pair tests. For example, Miyawaki et al (1975) tested NS of Japanese in their ability to discriminate the English /r/-/l/ contrast using a minimal pair test and found near chance performance. However, when the third formant (F3) of the stimuli was isolated (the primary acoustic difference between /r/ and /l/ is located in F3), such that the stimuli no longer sounded like speech, the NS of Japanese subjects performed as well as NS of English. Similarly, for both a Hindi (/d/-/ɖ/ - dental versus retroflex alveolar plosive) and Nthlkampx[3] (/q̰/-/k̰/ - glottalized uvular versus glottalized velar) contrast, NS of English were able to discriminate between isolated portions of initial syllables though they were unable to do so when the contrasts were presented in full syllables with attached vowels (Werker & Tees, 1984b). They were also able to discriminate the full syllables when the two stimuli to be discriminated were presented with a very short ISI (500ms). These experiments reveal that the lower level sensory mechanisms necessary for perceiving the sounds of language remain intact and therefore linguistic perceptual difficulties in an L2 are the result of this low-level information being discarded or ignored in perceiving. It seems to

[3] Note that in the initial paper (Werker & Tees, 1984b) this language is referred to as Thompson.

be the case that our perceptual system can in fact hear the required information. However, they have been trained over time to treat this as irrelevant variation.

It is of course not the case though that all novel L2 phonemes are certain to cause perceptual difficulties for L2 learners. In the Arabic research cited above, there were clear differences in the extent to which the pairs were learned. For example, the /ɣ/ - /x/ contrast, neither of which appears in English, was readily perceived by many relatively inexperienced listeners and nearly all of the advanced learners appeared to do as well as NS of Arabic. Best & Strange (1992) compared Japanese learners of English with NS of that language in discriminating and identifying synthetic stimuli featuring three different English approximant contrasts - /r/-/l/, /w/-/j/ and /w/-/r/. These pairs differed with respect to their status in the two languages. The /r/-/l/ contrast is a phonemic distinction in English, but not in Japanese, the /j/ is found in both languages, but its use is more restricted in Japanese since it is not permitted before high vowels in word initial position. Japanese and English /w/ differ in that English, unlike Japanese, round their lips when pronouncing it. They found no cross language differences in perceiving the /w/-/j/ contrast. For the /w/-/r/ contrast the two language groups performed similarly on end point stimuli, but exhibited different category boundaries – the Japanese subjects identified more stimuli in the continua as /r/ than the English subjects. NS of English perception of the /r/-/l/ contrast was clearly categorical, while for the Japanese it was nearly flat – they showed little evidence of an ability to tell the two apart. Polka (1995) compared discrimination of two German vowel contrasts by NS of English. They compared perception of German /ʏ/-/ʊ/ and /y/-/u/. All of these vowels are pronounced with rounded lips in German, and in both cases the difference is that the first of the pair is a front vowel and the second a back vowel. Since English does not have front rounded vowels, NS of English can also rely on the lip-rounding cue in making the distinction.

Despite the fact that the two pairs differ in the same feature, the NS of English were not significantly different than the NS of German in discriminating stimuli on a /y/-/u/ continuum, but did much worse on the /ʏ/-/ʊ/ continuum. Novelty alone then does not explain the differences in perceptual behavior by nonnative listeners. It has also been found that NS of English with no prior experience with Zulu perform as well as NS of Zulu in discriminating some of the so-called click sounds of that language, despite, or perhaps because of, the novel acoustic nature of the clicks for NS of English.

So, it is not the case that a given listener will fail to perceive any phonemic contrast not present in his/her L1, nor that all such contrasts will be equally difficult for L2 learners. In addition to large differences between individuals, the ability of a given group of L2 listeners to perceive a phonemic contrast not found in their language will be influenced by acoustic, phonetic and phonemic factors.

3.3.3 Acoustic, phonetic and phonemic factors

One possible factor in explaining the relative difficulty of acquiring L2 speech contrasts is acoustic salience. Denis Burnham (1986) proposed that speech sounds could be placed along a fragile/robust continuum based on the psychoacoustic salience of the sound and that robust sounds would require a longer period without exposure before the ability to perceive them would be lost. He claims that there are two distinct phases of perceptual loss. The early phase is during the first year of life and the second phase occurs between 4 and 8 years of age. Fragile contrasts are only maintained if relevant exposure occurs during this first phase, while robust contrasts are lost only if relevant exposure is lacking during the second phase. Furthermore, these robust contrasts are more easily recovered later in life. Burnham further proposes that it is not a coincidence that the second phase of loss coincides with the onset of formal linguistic training and that the learning of an alphabetic writing system further shapes phonemic perception. Although Burnham (1986) is frequently cited in the L2 speech learning literature there are few other experiments that test his proposals. In fact, the bulk of speech perception research has been conducted with either very young infants or with adults, while young children and juveniles have been more or less ignored (with the noted exception of research by Walley (e.g. 2007)) The contribution of acoustic salience to L2 learning difficulties remains unknown and as research described below will show, it is surely not the only factor. It should be noted though that the Arabic contrast that this dissertation is most concerned with (/h/-/ħ/) has very low acoustic salience as both members of the pair produce relatively little energy and take the formant structure of surrounding vowels.

Although acoustic salience may play some role in determining the difficulty with which an L2 contrasts will be acquired, differential success in acquiring different L2 phonemes is best explained with reference to the influence of L1 on L2 perception. In short, the evidence to be reviewed suggests that L2 sounds, at least initially, are perceived through the filter of the L1 speech perceptual system. Thus simple reference to L1 and L2 phonology is not sufficient for predicting the pattern of difficulties and more fine grained analyses of phonetic factors – differences in articulations and their acoustic consequences – is required.

The important influence phonetic factors is shown in a study nearly identical to the Strange & Best (1992) paper described above that performed the same series of tests on French, rather than Japanese listeners (Hallé, Best & Levitt, 1999). This experiment tested the French listeners on the same three continua used in the prior study – English /r/-/l/, /w/-/j/ and /w/- /r/. The experimenters note that unlike Japanese, French has phonemes that are considered phonologically parallel to English /r/-/l/, but that the articulatory and acoustic realizations of the phonemes are different in the two languages.

The French and English /l/ are relatively similar, though the English /l/ is more likely to be an alveolar articulation and the French a dental one. However the /r/s in the two languages differ considerably. The French /r/ is most frequently produced as an uvular approximant, sometime with frication while the English /r/ (or [ɹ] in a narrow transcription) is a palato-velar central approximant. Thus, although they are considered parallel phonologically they are in fact quite distinct phonetically. The experimenters wanted to know if these phonetic differences would alter the way the two groups perceived the continuum, or if the phonological status was all that mattered, such that their perception would be more or less identical. They found that the French speakers differed from NS of English in how they divided the /r/-/l/ continuum of English. Thus the phonological parallel between English and French /r/ was unimportant. The articulatory and acoustic differences between the two, perhaps not surprisingly, influence the way they are perceived.

The difference between phonetic and phonemic factors is illustrated by the conflicting results of the aforementioned Werker & Tees (1984b) study and another study (Pisoni, 1973) that also examined the effects of manipulating the interstimulus interval (ISI), that is, the period of time that elapses between the presentations of the individual stimuli to be compared. Pisoni (1973) examined the ability of NS of English to discriminate between two acoustically different members of the same category (/bæ/) and found that subjects performed poorly, regardless of the ISI. Werker & Tees (1984b) suggest that this is due to the fact that the difference between the paired stimuli in the Pisoni experiment was neither phonemic, nor phonetic. That is, the acoustic differences between the stimuli do not cross a boundary that is phonemic in any language. In the Werker & Tees experiment, the two stimuli do cross a phonetic boundary, though that boundary is not phonemic in English. Phonetic boundaries then are a function the physical nature of sound and the human (or even mammalian) auditory system and will be the same for any and all languages. Any given language will use a unique subset of these boundaries, which will be phonemic for that language.

The most comprehensive attempt to tease out the different factors bearing on the relative ease of discriminating an L2 contrast is provided by Polka (1991). In this experiment NS of English are tested on their ability to perceive the Hindi dental versus retroflex place of articulation across four different voicing contexts each with a different phonetic status for NS of English. The four were prevoiced (/d̪/-/ɖ/), voiceless unaspirated (/t̪/-/ʈ/), voiceless aspirated (/t̪^{h} /-/ʈ^h/), and breathy voiced (/d̪^{h}/-/ɖ^h/). Although none of these contrasts exist in English, some of them do exist as allophones in English. NS of English do have experience with retroflex stops as allophonic variants of alveolar stops. For example, the initial consonants of the words “try” and “dry” in English are pronounced as a retroflex [ʈ^h] and [ɖ^h] respectively. The prevoiced retroflex /ɖ/ may exist in “address”, and dental /t̪/ and /d̪/ also exist in words like “eighth” and “width”. Taken together this means that NS of English will have linguistic experience with both phones in the prevoiced and voiceless aspirated contrasts (/d̪/-/ɖ/ and /t̪/-/ʈ/), only one of the phones in the voiceless aspirated context(/t̪ h/-/ʈ h/), and neither of the phones in the breathy voiced contrast(/d̪ h/-/ɖ h/). Polka reasons that if phonemic status is the only relevant factor in determining the difficulty of perceiving an L2 contrast, then each of the four contrasts should be equally difficult for NS of English since none of them exist in English. If, however, phonetic factors are also important, then the familiarity of subjects with both phones of the voiceless unaspirated and the prevoiced pairs should make those the easiest of the four while the breathy voiced contrasts, in which neither phone appears, should be the most difficult. The voiceless aspirated pair would be expected to be intermediate in difficulty. Polka also argues that the voiceless aspirated contrast is the least acoustically salient one, due to a lack of clear formant structure caused by the aspiration, and so if acoustic factors are most important, as argued by Burnham (1986) above, this should be the most difficult contrast for the NS of English. The stimuli used in the experiment were CV syllables recorded from a NS of Hindi. The subjects' task was AX discrimination. Four different subject groups were used, one for each of the four contrasts. The results of the test indicate that there were significant differences in subject performance, which suggests that phonemic factors alone do not explain the difficulty faced by L2 learners in perceiving foreign language phonemes. However, the results also fail to find strong evidence for the importance of phonetic familiarity since the prevoiced contrast, in which NS of English could be expected to be familiar with both phones, was the most difficult, and the only contrast for which NS of English did not perform significantly better than chance on. In regard to acoustic factors, the breathy voiced pair was not shown to be more difficult than the other pairs and so it seems that acoustic factors are also not overwhelming (or that this operalization of acoustic salience is inappropriate). It should also be noted that different subjects were used for each of the four contrasts, and there was substantial variation between subjects, such that it is difficult to draw strong conclusions from the research. Nevertheless, it is one of the few comprehensive attempts to measure the relative contributions of phonetic, phonemic and acoustic factors.

The influence of phonetic factors is also explored in another experiment by Polka (1992). In her first experiment, she compares NS of English and NS of Persian in their ability to discriminate the same Nthlkampx contrast (/q̰/-/k̰/ - glottalized uvular versus glottalized velar). Polka reasoned that if native language experience with a given place contrast facilitates perception of a similar place contrast in a foreign language, then NS of Persian, which contains a phonemic contrast between velar and uvular place of articulation in a different voicing and manner than the Nthlkampx contrast, will outperform NS of English who have no experience with uvular articulations. She found though that, if anything, the Persian NS performed more poorly on the task. Therefore, even though Persian listeners have experience with a velar-uvular contrast it did not help them perceive a similar contrast in another language.

In a second experiment, Polka (1992) compared the performance of NS of English on contrasts from two languages – the same Nthlkampx contrast and the Persian velar versus uvular stop contrast (/g/-/ɢ/). Since English contains one of the Persian phonemes (/g/), but neither of the Nthlkampx ones, Polka reasoned that if phonetic similarity plays a role then NS of English should perform better on the Persian contrast. Although the final results were somewhat confused by significant order effects (the subjects performed much better on the contrast that they were tested on first, regardless of which one it was), she did not find facilitative effects for phonetic familiarity as the NS of English performed equally poorly on both contrasts.

It has also been suggested that the acoustic correlates of some articulations may be so novel to certain nonnative listeners that they are not processed as speech signals and therefore are easily discriminated. Best, McRoberts & Sithole (1988) found that NS of English were not significantly worse than native speakers of Zulu in discriminating some Zulu click consonants. A subsequent study (Best, Traill, Carter, Harrison & Faber, 2003) compared NS of English with NS of Isizulu and Sesotho (both click languages) on the perception of bilabial clicks of the !Xóõ language and found that while the NS of English performed well, those of the other click languages had considerably more difficulty, which the authors argue was due to assimilation of the !Xóõ language sounds to phonemes of their own language while for NS of English that were not perceived as linguistic at all.

These experiments suggest that in some cases when the phonetic properties of the native language are similar, but not identical, to those of the second language that this familiarity confounds the perception of L2 sounds. A possible mechanism explaining this effect is proposed by Patricia Kuhl's Native Language Magnet (NLM) model of speech perception, which is discussed below.

Differences in native and nonnative processing of speech sounds have also been investigated with neuroimaging techniques. The mismatch-negativity (MMN) is an event related potential (ERP) component that has been used to explore differential processing of speech sounds by native and nonnative listeners. An ERP is a comparison of electrical activity measured at the scalp in response to one event (e.g. exposure to a word) compared to the response to a different event (e.g. exposure to a nonword). Characteristic response patterns are referred to as components. The mismatch negativity, which occurs in response to a change in any constantly repeated acoustic signal occurs within 100-200ms after the presentation of the deviant stimulus (see Näätänen, 1995 for a general review and Näätänen, 2001 for a speech specific review). Although the MMN occurs in response to any change in an auditory stimulus, it has been found to be qualitatively and quantitatively different in response to an acoustic change that crosses a phonemic boundary. Näätänen et al (1997) compared MMN responses to changes in repeated vowel sounds in both Finnish and Estonian listeners. When the change stimulus crossed a boundary that was phonemic for Estonian, but not Finnish, listeners the MMN was more intense in the Estonian listeners and was more localized to the left-hemisphere. This differential response develops in infants somewhere between 6 and 12 months of age (Cheour, Ceponiene, Lehtokoski, Luuk, Alli, Alho & Näätänen, 1998), but can also be evidenced by late learners of a foreign language, which suggests that L2 processing of phonemes can, at least in some circumstances, show evidence of nativelike processing.

Winkler, Kujala, Titinen, Sivonen, Alku, Lehtokoski, Czigler, Csépe, Ilmoniemi & Näätänen (1999) compared NS of Hungarian that had no experience with Finnish and those that were advanced late-learners of that language on a vowel change that is phonemic in Finnish, but not Hungarian. While the NS of Finnish and the NS of Hungarian who were highly proficient in Finnish showed robust MMNs to the stimulus change, the naive NS of Hungarian showed no MMN response. Tremblay, Kraus, Carrell & McGee (1997) trained a groups of NS of English to perceive a bilabial VOT contrast that is not phonemic in English and found that the MMN response to deviant stimuli occurred prior to improvement on behavioral measures of discrimination. It seems that first lower level processing must respond differentially to the stimuli before higher levels can notice the change and modify behavior. The improvement in the MMN was largest in the left-hemisphere and also transferred to a similar VOT contrast for an alveolar place of articulation.

Näätänen et al (1997) argue that the MMN is a reflection of low-level auditory processing and therefore, since it shows language specific effects, language learning must influence low-level sensory processing, supporting the NLM theory of Kuhl and others (see below); however, it has also been argued that the MMN reflects higher level processing and so cross-linguistic results are not indicative of effects of language experience on low-level processing (Sharma & Dorman, 2000).

3.3.4 Phonetic context and L2 speech perception

One of the most important conclusions drawn by researchers of second language speech perception is that the appropriate level of analysis in exploring nonnative speaker segmental perception is the context conditioned, allophonic level. As we have seen, familiarity with the feature of voicing does not necessarily mean that NS of English can detect voicing changes in all circumstances (Polka, 1991), thus the phonetic feature level is inappropriate. By the same token, we cannot assume that the ability to discriminate between allophones of a given phoneme in one context will necessarily transfer to discrimination between allophones of the same phoneme in a different context. This has been show most clearly in NS of Japanese perception of the English /r/-/l/ contrast. Sheldon & Strange (1982) gave NS of Japanese a minimal pair test of English words that contrasts English /r/ and /l/ in a variety of different positions within a word – word initial, consonant cluster, intervocalic and word final. They found that the Japanese subjects had the most difficulty with the contrast in prevocalic consonant clusters (e.g. “clam vs. “cram”) and that they performed as well as NS of English in the word final position. An earlier attempt to explain this reasoned that increased duration of the production of /r/ and /l/ facilitated the task by providing more information (Dissoway–Huff, Port & Pisoni, 1982), however, it was subsequently found that NS of Cantonese, who also show difficulty in acquiring the English /r/-/l/ contrast, showed a different pattern of results. These subjects were best at the discrimination in word initial and medial positions and were worst in word final position. This further confirms that native language speech perception has a powerful influence on how L2 phonemes are perceived. Thus, while language teachers often refer to difficulties that their learners have with a given phoneme, this is generally speaking an insufficient level of analysis. Indeed we shall see in Chapter 5 that NS of English are significantly worse at discriminating Arabic /h/ and /ħ/ in the word initial and intervocalic position than in other phonetic contexts. Therefore, it is important to consider phonetic context when diagnosing difficulties faced by L2 learners. As we shall see in the next chapter, it is also the case that training in one context does not generally transfer to gains in other contexts (e.g. Lively, Logan & Pisoni, 1993).

As noted in the prior chapter, the phonotactic rules of the native language also influence how L2 speech sounds are perceived and it appears that listeners attempt to impose the same constraints on all speech sounds. Recall the experiment by Massaro & Cohen (1983) in which the phonotactic permissibility of a sequence affected perception. A synthetic sound falling between /r/ and /l/ was more likely to be perceived by NS of English as /r/ when preceded by /t/, but as /l/ when preceded by /s/ presumably because /tl/ and /sr/ are impermissible word initial sequences. This fact influences the perception of second language speech sounds. For example, Peperkamp, Pettinato & Dupoux (2003) tested NS of French in their ability to discriminate between the uvular voiced fricative [ʁ] and the uvular voiceless fricative [χ]. The latter is found only adjacent to voiceless consonants in French and the experimenters were interested in how presentation context would effect perception of the distinction. In one experiment they presented subjects with simple VC monosyllables ([aʁ - aχ]) and in another with VC.CV syllables some of which were phonotactically permissible in French and others that violated the expected spreading of voicelessness. It was found that while the subjects were able to discriminate the contrast in isolated CV syllables, they appeared to lose that ability when the contrast was contextualized. The authors of the paper speculate that the decontextualized forms may have allowed subjects to focus more on acoustic factors while the contextualized forms forced them to rely on linguistic analysis, which ignored the differences between the two, essentially treating them as belonging to the same category. Similarly, NS of Mandarin Chinese, which prohibits word final voicing, have difficulty discriminating between /d/ and /t/ when they occur word finally, but not word initially (Flege, 1995b) which is presumably due to the fact that Mandarin does not allow voicing on final consonants.

These phonotactic constraints not only cause nonnative listeners to misperceive certain sounds, but also to perceive sounds that do not exist in the speech stream. Dupuox, Kakehi, Hirose, Pallier & Mehler (1999) found that NS of Japanese, unlike NS of French, report perception of the vowel /u/ in the nonword /ebuzo/ even when it has been completely removed from the acoustic signal. Furthermore, they have great difficulty discriminating between the nonwords /ebzo/ and /ebuzo/ in an ABX task even when the X stimulus is acoustically identical to one of the first two stimuli (French listeners had no such difficulty). Presumably this is due to the fact that in Japanese, consonant clusters (except those with nasals) are impermissible. In a follow up study using brain imaging technology (ERP) it was found that the speech perceptual system incorporates phonotactic information very early in the word identification process (Dehaene-Lambertz,Dupoux & Gout, 2000). Subjects were presented with 4 stimuli from the same category (/ebzo/ as spoken by 4 different individuals) followed by a deviant stimulus (/ebuzo/). The French, but not the Japanese subjects showed an MMN-like response to the deviant stimulus in the 139-283ms window. They conclude that this result supports speech perception models in which the syllable, rather than the phoneme is the relevant unit. However, they also note that the Dupoux et al study above found that although the Japanese were far worse than the French in detecting the difference between /ebzo/ and /ebuzo/ they did show some ability to do so, and so additional factors may be at work.

3.3.5 Acoustic Cue Reliance in L2 Speech Perception

It has also been found that nonnative listeners frequently rely on different acoustic cues than native speakers in making phonemic judgments and frequently rely on fewer cues than NS. Iverson, Kuhl, Akahane-Yamada, Diesch, Tohkura, Ketterman and Siebert

(2003) compared NS of English with NS of Japanese in the cues that they rely on in discriminating between /r/ and /l/. They used synthetic stimuli of /ra/ and /la/ tokens that varied in both F3 and F2. Subjects were presented with two stimuli and asked to rate the similarity between the two on a scale of 1-7. They found that while NS of English relied almost exclusively on F3, NS of Japanese relied more predominantly on F2, which is not relevant to the discrimination. Flege & Hillenbrand (1986) compared NS of English, French, Finnish and Swedish in the use of length cues in discriminating between word final /s/ and /z/. For NS of English both the length of the preceding vowel and the length of the final consonant frication are used to discriminate between them. Thus as the vowel becomes shorter NS of English are more likely to perceive /s/. As the frication portion becomes shorter, NS of English are more likely to perceive /z/. French is similar, and so that group served as an additional control. Neither Swedish nor Finnish has a word final /s/-/z/ contrast, however Swedish is similar to English and French in that word final voicing is perceived with reference to the length of the frication and the preceding vowel. In Finnish vowel and quantity length are phonological and it is possible to have a long vowel followed by long frication as well as a short vowel followed by short frication. They found that both the NS of English and French integrated vowel and frication length cues in making the discrimination, but both the NS of Swedish and of Finnish relied almost exclusively on the vowel duration cue. Again, we see language learners borrowing perceptual strategies from their native language in attempting to perceive a new language. Flege et al (1997) provides evidence that L2 learners can modify their perceptual strategies to become more nativelike. As noted, the more experienced learners of English had come to rely more on quality than quantity information in vowel perception in contrast to their less-experience counterparts.

The observation that L2 listeners frequently rely on fewer cues than NS speakers has led researchers to wonder if L2 listeners can learn to integrate multiple cues and if so what factors facilitate that process. Underbakke, Polka, Gottfried & Strange (1988) report that both skilled and unskilled NS of Japanese demonstrated reliance on both temporal and spectral parameters of the English /r/-/l/ distinction and that more skilled learners did so more reliably, some as well as NS of English. However, a study by Ingvalson, McClelland & Holt (2011), despite finding significantly more accurate identification of /r/ and /l/ by NS of Japanese with longer terms of residence, found no evidence that they had increased reliance on F3 in making the discrimination. The authors speculate that the improvements are the result of more efficient use of L1 categories in classifying L2 speech sounds and not due to the formation of new categories.

Several training studies have also addressed the issue of cue weighting for L2 listeners and have found evidence that L2 learners can learn new cue weightings in L2 speech perception and that they can even attend to new, previously ignored dimensions (Holt & Lotto, 2006; Francis, Baldwin & Nussbaum, 2000; Francis & Nussbaum, 2002). We shall examine these studies in more detail in the next chapter.

3.4 The link between segmental perception and production in an L2

It is also of interest to this dissertation to probe the relationship between the perception of L2 segments and the production of those segments. It has been proposed (Flege, 1995a; Rochet, 1995, and see discussion of SLM below) that inaccuracies in L2 speech production sometimes have a perceptual basis. That is, the L1 phonology filters out important auditory information such that the L2 speaker does not have an accurate model to mimic. In circumstances in which this is the case we would expect perceptual improvements to lead automatically to productive improvements.

For example, Flege (1993) examined the perception and production of word final /t/ and /d/ in English among NS of Taiwanese and Mandarin, neither of which have a /t/ contrast in the word final position. NS of English use the length of the preceding vowel as a cue to word final stop voicing and typically produce longer vowels before /d/ than /t/. They found that L2 listeners that relied on vowel length as a cue to identification also showed greater differences in their vowel length productions. The extensive study by Flege et al (1997) compared perception and pronunciation of American English vowels by NS of four different languages (Spanish, Mandarin, German and Korean) with varying experience with English as an L2. Acoustic measurements and NS ratings of the subjects' productions of /i/, /ɪ/, /ɛ/ and /æ/ were compared with their identification of stimuli along continua of vowels from /i/ to /ɪ/ and from / ɛ/ to /æ/ and it was found that perceptual performance accounted for a significant amount of the variability in production, but also that much variation remained unexplained. Hattori and Iverson (2010) looked for correlations in the perception and production of English /r/ and /l/ among a group of 46 NS of Japanese in England. Subjects performed identification and discrimination tasks with synthetic stimuli varying in multiple dimensions (most notably F2 and F3) and also provided speech samples, which were rated by NS and acoustically analyzed. They found a moderate correlation between subjects' identification accuracy and the intelligibility of the pronunciation. However, pronunciation did not appear to correlate with improved sensitivity to changes in F2 and F3 and conclude that perceptual and productive knowledge are not necessarily linked.

It should not be assumed that an individual that cannot perceive a given phoneme cannot accurately pronounce that phoneme. For example both Goto (1971) and Sheldon

& Strange (1982) found NS of Japanese that performed poorly in tests of their ability to perceive the /r/-/l/ contrasts, but whose productions of those phonemes were accurately identified by NS of English. Presumably, these individuals lack the ability to monitor their own productions, but have nevertheless figured out that certain articulations are less likely to lead to misperception by listeners than others.

Training studies, which will be reviewed more thoroughly in the next chapter, have also found that perceptual training can yield improvements in pronunciation (Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1997; Rochet, 1995; Hardison, 2003; Hazan, Sennema, Iba and Faulkner, 2005; Wang, Jongman & Soreno, 2003). However, as with the experiments described here, there was substantial variability. For example, in the Bradlow et al study, there were examples of individuals that showed perceptual improvement, but no productive improvement and vice versa.

3.5 Models of second language speech learning

Many of the phonetic training studies to be reviewed in the next chapter were conducted to test the claims of one or another model of L2 speech perception. In what follows the most important of these models and the research findings that they rely upon will be briefly reviewed.

3.5.1 The Speech Learning Model (SLM)

Perhaps the most influential model of L2 speech perception is the Speech Learning Model of James Flege (most explicitly described in Flege, 1995a, but see also Flege, 1992; 1999; 2002; Flege, Schirru & MacKay, 2003). From the point of view of a foreign language educator, the most important claim of the SLM is its rejection of the Critical Period Hypothesis (CPH) for speech learning. The CPH proposes that at some point in development there is a qualitative change in the way that language is acquired such that learners who are exposed to the L2 after the critical period closes will not be able to reach native like proficiency in the language (see Birdsong, 1999 for a review). It has also been claimed that because speech production and perception are directly linked to complex articulatory movements that the effects of the CP are likely to be particularly evident in the realm of speech production (Long, 1990). The SLM, however, proposes that “the mechanisms and processes used in learning the L1 sound system . . . remain intact over the life span.” (Flege, 1995: 239) Flege points to the results of the experiment described above (Flege et al, 1995) which showed that age of arrival accounted for a significant amount of the variation in accentedness ratings for 240 subjects and no discontinuity in the graph plotting AoA against accentedness ratings. In other words, there was no evidence for a CHP. This study was repeated with NS of Korean and the AoA affect held up even when other confounding factors were controlled for (Flege et al, 1999). Thus, the claim that for accurate segmental production ‘earlier is better’ was confirmed. This is not; however, the same as a critical period. If a critical period for speech learning exists, we would expect that all subjects arriving after the close of that critical period would be equally poor in producing target like speech and all those before would be nativelike. However, the data provided by Flege and his colleagues suggest that ‘earlier is better’ remains true throughout the lifespan and there is no evidence for a qualitative change in how speech sounds are acquired. The SLM proposes then that the difficulties experienced by second language learners are not due to the passing of some critical period, but rather due to the existence of a fully developed L1 speech perceptual system which influences the way that all speech sounds, L1 or L2, are perceived and produced.

The SLM rests upon 4 main assumptions about second language speech learning (see Flege, 1995:239). First, as noted above, that the same mechanisms employed in learning the L1 speech system is available to adult learners learning L2 speech sounds. Second, that the relevant unit in segmental speech perception is the phonetic category – abstract, position specific speech sounds that exist somehow in the brain as memory representations. Although these categories are less abstract than phonemes, they are still abstract in that they are the representations that are obtained after normalization filters out linguistically irrelevant details. Third, the categories established in childhood for L1 can evolve over time to reflect the properties of all speech sounds, L1 or L2, that are identified as category members. And fourth that L1 and L2 speech sounds share a common phonological space and bilinguals must maintain a contrast between the two.

Based on these assumptions, several hypotheses follow, the most relevant of which are mentioned here. If learners are able to perceive some phonetic difference between an L1 and L2 speech sound, then a new category for the L2 sound may form and the greater the perceived dissimilarity between the two sounds, the more likely a new category will form. As AOL increases, the ability of discerning these phonetic differences, and of forming new categories, will decrease. When category formation is blocked, a single category is used to process both L1 and L2 sounds and the category will come to represent a compromise between the two sounds. A newly formed L2 category will not necessarily come to be identical to the same category in monolingual speakers of the L2 because the new category may be "deflected" away from a previously existing L1 category. Finally, the production of sounds will eventually correspond to their perceptual specification.

The effects of equivalence classification are illustrated by Flege (1987). He compared the voice onset time (VOT) used by monolingual NS of English with bilingual French-English speakers that were also NS of English. Voice onset time is a measure of the time between the release of the consonant and the start of vocal cord vibration.

Monolingual NS of English typically produce /t/ with a VOT longer than that of monolingual NS of French. Flege found that the bilingual subjects pronounced /t/ with a VOT shorter than typical in English, but longer than typical in French. Thus, having equated French and English /t/, these subjects /t/ category had come to reflect the average VOT of all the /t/s that they heard, in either French or English. Similarly, bilingual NS of French also produced /t/s with compromise VOT values. Thus, learning a foreign language can change how the native language is perceived and produced. A similar study with NS of Spanish that began learning English in childhood (5-6 years of age) showed that these subjects maintained a distinction between Spanish and English /t/ (Flege, 1991) and Flege argues this may be because an L1 category had not already been firmly established when the subjects began to be immersed in English, so they were able to form separate categories for Spanish and English. This phenomenon may also explain the results of the MacKay et al (2001) study described earlier. In that study, NS of Italian that used their L1 infrequently were more native-like in perceiving VOT contrasts than AoA matched counterparts that still used the L1 a lot. Without exposure to Italian-like VOT contrasts, these subjects' categories had become increasingly like those of monolingual NS of English.

The SLM proposes that the relative difficulty of learning a new foreign language sound will be dependent on the perceived phonetic dissimilarity between that sound and the closest L1 equivalent. In practice this claim, and similar ones like it from other models of speech perception, has been difficult to test empirically as there is no principled way to measure the phonetic similarity of speech sounds. Flege (1992) himself has suggested several possible ways to operationalize the notion of phonetic similarity including using IPA symbols as proxy measures or using rhyming judgments or dissimilarity ratings, but none has proven workable thus far (but see Strange, 2007 for a comparison of methods for assessing cross-language similarity of vowels). The only currently accepted method for determining the perceived phonetic similarity between a L1 and L2 speech sound is through ratings of L2 sounds by the L1 group in question. For example, Guion et al (2000) asked NS of Japanese to identify different English consonants in terms of their own sounds and then to rate, on a scale of 1-7, how close the L2 sound was to the identified L1 sound. They found that English /b/, /s/, /t/ and /v/ were all rated as good fits to their Japanese counterparts, while /r/, /l/ and /θ/ were poor fits. They subsequently compared the performance of three different groups of NS of Japanese with varying levels of English language experience in a categorial ABX discrimination task with a number of different English phoneme contrasts and also English-Japanese contrasts (e.g. English [ɹ] versus Japanese [ɾ]). They found that, contrary to the predictions of the SLM, only one of the poor fitting contrasts (the [ɹ] - [ɾ]) showed evidence of learning while two others /θ/-/s/ and English /l/-/r/ showed no such evidence. They conclude that "the SLM cannot be readily extended to early stages of L2 speech acquisition without further investigation (Guion et al, 2000: 2723).

3.5.2 The Perceptual Assimilation Model (PAM)

Another influential model of second language speech perception that has motivated research in phonetic training is the Perceptual Assimilation Model (PAM) developed by Catherine Best (1995). It should be noted at the outset that PAM and SLM are not directly comparable. While both are concerned with explaining second language speech perception, SLM makes predictions about the end state of L2 speech learning, after long immersion, while PAM makes predictions about perception of L2 sounds when first encountered by nonnative speakers.

PAM is based in general upon direct realist theories of perception (see Gibson, 1988) and more specifically on gestural theories of speech perception (see Fowler & Galuntucci, 2008). Briefly, gestural theories of speech perception propose that the objects of perception are not the auditory patterns of speech, but rather the articulations that cause those patterns. Proponents of gestural theories of speech perception rely on a number of different types of evidence. As mentioned in the previous chapter, the acoustic correlates of /d/ in the syllables /du/ and /di/ are, despite the common articulation, very different, and yet yield the same precept in a listener. Proponents of gesture-based accounts argue that this would only be possible in the event that perception was somehow mediated by reference to articulation. However, it has been shown that quail (Coturnix coturnix), trained to perceive /d/ (and other stop consonants) in one set of vowel contexts, successfully generalized that learning to new vowel contexts, despite the acoustic differences between the two (Kluender, Diehl & Killeen, 1987). It is inconceivable that quail did this through reference to human articulatory gestures. It is also argued that the McGurk effect, described in the previous chapter, is further evidence of reference to articulation, though others argue that is merely evidence of the fact that listeners have learned to take advantage of all perceptual information when learning to perceive speech sounds (Diehl, Lotto & Holt, 2004). While most of the evidence cited in support of the gestural account can be explained more parsimoniously in other ways, a experiment by Fowler & Dekle (1991) is more convincing. They found a haptic equivalent of the McGurk effect. Some subjects' perception of spoken syllables was influenced by lip movements detected by placing their fingers over the lips of speakers. Fowler & Dekle report that at least some subjects integrated the haptic and auditory information in perception. Since the subjects had no previous experience with Tadoma (a method of reading lips through touch), a learning explanation fails to explain the result.

It should be noted that data from almost half of the subjects of this experiment was discarded, and despite the importance of the result to the gestural account the experiment has either not been repeated, or repetitions of the experiment have failed to yield publishable results. At present, the gestural account remains speculative.

Thus, PAM takes articulations as the perceptual primitives in speech perception. This means that a listener gathers data (i.e. acoustic visual and perhaps even haptic) and analyzes that information to make a best guess at what articulations were made by the other speaker and then uses that information to reconstruct the intended message. PAM proposes that the relative difficulty an L2 listener will have in perceiving a given L2 phone relates to the articulatory similarity of that phone with a native language one. PAM delineates a number of different possibilities for how two foreign language phones might be perceived by a nonnative listener on first encountering the sounds:

1. In two-category (TC) assimilation both of the nonnative phones will assimilate to a different native language category. In this case nonnative listeners will have no difficulty discriminating between the two.
2. For a category goodness difference (CG) both members of the pair assimilate to the same native language phone, but one does so better than another. Depending on the magnitude of the discrepancy, discrimination is expected to range from moderate to good. For NS of English learning Arabic it may be that this is the case for ka:f (/k/) and qa:f (/q/). While both are frequently perceived as /k/ by our early learners the qa:f is a poor fit and the discrimination appears to be readily learnable (Burnham, 2010).
3. In single category (SC) assimilation both phones assimilate to the same native phone and are equally good, or equally bad, exemplars of the category. In this event, discrimination is expected to be poor. These will cause the most difficulty

for would be learners of a foreign language. For NS of English, this appears to be the case for a number of Arabic phoneme pairs (Burnham, 2010), including the pair that is the focus of our training study – the voiceless glottal fricative /h/ vs. the voiceless pharyngeal fricative /ħ/.

4. For the uncategorizable type (UU) both phones fall within the phonetic space of a listener, but neither is categorized as a native speech sound. It is difficult to predict discrimination in this event. This may be the case for our learners in discriminating the voiceless uvular fricative /χ/ from the voiced uvular fricative /ɣ/, which appears to be relatively easily learned by NS of English (Burnham, 2010).
5. In the uncategorized versus categorized type (UC) one phone assimilates to a native phone, while the other does not. This should result in good discrimination.
6. Finally, in the nonassimalable type (NA) neither of the phones is perceived as a speech sound by the listener. In this event, it is again difficult to predict discriminability. This may be the case when NS of English are exposed to the click sounds of some African languages (see Best et al, 1988).

To give an example of how these claims are tested, consider an experiment by Best, McRoberts and Goodell (2001) that tested NS of English (with no exposure to Zulu) on their categorial discrimination of three different Zulu contrasts. The three contrasts were a voiceless versus a voiced lateral fricative (/ɬ/-/ɮ/) a voiceless aspirated versus and ejective velar stop (/kʰ/-/k'/) and a plosive versus implosive bilabial stop (/b/-/ɓ/). Based on prior assumptions and on subjects transcriptions and descriptions of each of the phonemes, they assumed that the /ɬ/-/ɮ/ contrast would be a two category (TC) contrast, that /kʰ/-/k'/ would be a category goodness contrast (CG) and that the /b/-/ɓ/ would be a single category (SC) contrast. Subjects performed a categorial discrimination task on single syllable stimuli and it was found that generally speaking the expectations of PAM were confirmed – the TC contrasts was easier than the CG contrasts which in turn was easier than the SC contrast. However, there were some discrepancies among individual subjects. For one, some subjects did much better than chance on the SC contrast while others did poorly on the TC contrast. Furthermore, some subjects perceived the /b/-/ɓ/ as a TC contrast, and yet were no better than those that perceived it as an SC contrast. The experimenters hypothesize that task order effects may explain the discrepancy.

The Guion et al (2000) study above is one of the few that directly compares predictions of the two models to determine if either or both can be extended (recall that the SLM is concerned with endstate learners and PAM with naïve learners) to predict performance in naïve (for the SLM) or experienced (for PAM) learners. As noted, the SLM does not well account for the findings. The PAM though accurately predicted the failure of the learners to acquire the English /l/-/r/ contrast as the Japanese ratings of those two phones indicated that both were uncategorizable (UU) to Japanese phonology and in between two existing categories, for which the PAM predicts learning difficulties.

3.6 CONCLUSION

The research reviewed above provides valuable information for the foreign language educator that will be useful in helping students to reach desired levels of proficiency. As noted in Chapter 1, segmental perception underlies a number of language skills important for proficiency and other things being equal a student that can accurately perceive L2 sounds is likely to be more capable than one that cannot.

As indicated in the first part of this chapter a number of factors affecting the outcome of L2 speech learning relate to learning context. In general, successful outcomes are aided by an early start, quality experience and abandonment of the L1. Of course, a university program has little control over the age at which their learners start learning, and while in such a setting it is unreasonable to imagine that students will abandon their L1, the importance of quality and meaningful use of L2 (e.g. Flege et al, 1997) in achieving positive outcomes should be noted, and even communicated to students.

The most important conclusion regarding linguistic factors that influence success is that the phoneme is not an appropriate level of analysis in diagnosing perceptual difficulties faced by L2 learners. A number of the studies cited above (e.g. Sheldon & Strange, 1982; Peperkamp et al, 2003; Flege & Wang, 1989) have shown that phonetic context influences the relative difficulty with which L2 speech sounds are perceived. These context effects can relate either to position in a word, or the specific speech segments surrounding them.

The models of speech perception reviewed, if not always explicitly, adopt the idea of "full transfer" and assume that L2 learners engage the entirety of their L1 perceptual system in decoding L2 speech sounds. All of the models also accept that whatever difficulties may exist in initially encountering L2 sounds, they can at least be partially overcome and improvements can be had in perceiving even the most difficult of sounds. It is the task of the language teacher then to facilitate the process by which this learning occurs. One intriguing means by which this may be done is called phonetic training and has shown to be a promising method for improving low-level perceptual skills in short periods of time. In the next chapter we define phonetic training, describe its methodologies and attempt to discern best practices from published studies using phonetic training in a variety of different L1/L2 contexts.

Chapter 4: Phonetic Training

4.1 Introduction

The purpose of this chapter is to describe in detail what phonetic training (PT) is and how it can be used to help us meet our pedagogical goals for Arabic. The chapter consists of two main sections. In the first, I describe the early history of PT experiments, which were designed to test many of the theories and claims of research described in the previous two chapters. This first section concludes with a description of two seminal studies in PT – Strange & Dittman (1984) and Logan, Lively & Pisoni (1991). The former of these studies was the first to combine this early PT research with research conducted in the field of second language speech perception and established the most important criteria for judging the success of a PT study in inducing speech learning – successful generalization of training gains to real-speech stimuli to which subjects have not previously been exposed. The latter established the specific methodology that has since been used to train a diverse array of speech contrasts in a variety of different L1/L2 combinations. In the second section of the chapter, I survey the important findings of PT studies since the Logan et al (1991) study with the primary goal of determining the requisite components of a methodologically sound, effective and informative PT study.

4.2 History of phonetic training

For the purposes of this chapter, we will define phonetic training as an effort to change patterns of speech perception in learners through relatively short-term laboratory training methods. By short-term we mean that total time on task will be measured in minutes or hours and occur over a time period of no more than several months. In general, phonetic training studies involve subjects performing perceptual tasks that are delivered via a computer terminal. In most cases, the metalinguistic knowledge provided to subjects is minimal or nonexistent. In other words, subjects are generally not told how different sounds are articulated, or how the acoustic consequences of those articulations may be described. Also, in general, subjects are given identical pre- and post-training tests designed to measure the effect of training. Although many training studies note the potential value of phonetic training as a tool for second language learning, the explicit research goals are focused on questions related to second language acquisition and speech perception, not language pedagogy.

The earliest known phonetic training study is a 1962 paper by Lane and Moore. The experiment had only a single subject, an aphasic patient that had lost the ability to discriminate between the phonemes /t/ and /d/ due to brain trauma. Land & Moore (1962) used synthetic stimuli along a VOT continuum from /do/ to /to/. In a pretesting identification task, the subject identified nearly all of the tokens as /do/. In the training phase, the subject was presented with only the endpoints of the continuum and asked to identify them as either /do/ or /to/. There were a total of 112 stimuli, evenly divided between /do/ and /to/, and the subject received immediate feedback on his response. Following training, the subject's performance on identification and discrimination tasks using the same stimuli was much more similar to that of a normal hearing control subject. Not only did he label the continuum much the same as a normal subject, but he also showed a peak in discrimination ability at the /do/-/to/ border. Thus, this first (published) effort at phonetic training was successful in the reacquisition of a lost ability to discriminate between these stimuli. The researchers do not tell us if the subject's newfound ability transferred to the real world of speech communication.

Most of the earliest phonetic training studies were inspired by research conducted at the Haskins lab in the 1950s and 60s, some of which has been reviewed in Chapter 2. This early research introduced many of the peculiar, but now quite familiar, phenomena of speech perception. Recall that in his Motor Theory of speech perception, Liberman claims that humans perceive sounds in either an acoustic mode, used for processing all nonspeech sounds, or a speech mode, which is exclusive to speech sounds from the earliest levels of processing, and that these two modes were at times in competition with one another (Liberman, 1996). For example, in one study (Liberman, Harris, Kinney and Lane, 1961) it was shown that subjects perceived speech stimuli categorically, but perceived nonspeech control stimuli continuously. In arguing against a special mode for speech perception, Lane (1965) presents the results of an unpublished laboratory study (Lane & Schneider, 1963) that used the same nonspeech stimuli as Liberman et al (1961) in a training study. Subjects were trained to identify endpoints of the continuum with immediate feedback. Three of eleven subjects were selected for further training. In posttest tasks, the trainees showed patterns of identification and discrimination similar to those for speech sounds, that is, poor within category discrimination and excellent between category discrimination. The results suggest it is not the special speech mode that imposes categorical perception, but rather the behavioral relevance of categorization that influences the development of perceptual patterns. In another unpublished study by Cross & Lane (1962), which is described in Lane (1967), subjects were trained to hum in one of two different frequencies in response to tones at two different decibel levels. In later exposure to tones along a decibel level continuum, subjects labeled the continuum categorically, as they were trained, and not continuously, as they could have done by varying the frequency level of their humming according to the decibel level of the tone. Lane (1967) also cites evidence that categorical perception can be induced in visual perception and in animal perception as well, and therefore there is no reason to assume a special mechanism for speech perception.

Lane's claim that patterns of speech perception could be easily manipulated in the short-term with simple laboratory procedures inspired further investigations of the results of such training procedures. Pisoni (1971) used the Lane & Schneider training paradigm to train six subjects on a continuum of isolated F2 transitions, which are perceived as nonspeech signals. Subjects were given both identification and discrimination training on the continuum, after which subjects were, on average, more consistent in identifying the stimuli and better at discriminating between them as well. However, Pisoni notes that there was a lack of any consistent training effect across subjects and large individual differences between subjects. He concludes that the selection of subjects may have a large impact on subsequent results. This is particularly relevant as both Lane and Liberman and colleagues routinely prescreened subjects and used only the subset most likely to support their own conclusions.

In a further test of Lane's contention, Strange (1972) conducted a series of training tasks intended to test the plasticity of speech perceptual processes. Strange used several different training tasks – discrimination, identification and scaling, in which the subject attempts to rate how far away, acoustically speaking, a given stimulus is from a standard stimulus. She reasons that "the nature of the processes underlying the categorical perception phenomenon and the effectiveness of various training procedures in modifying adults' perception are intimately related." (Strange, 1972:19). Strange trained subjects on a VOT continuum and wanted to determine if NS of English could come to perceive Thai prevoicing. Strange had a small number of subjects for each experiment and there was once again substantial diversity in subject responses to training. Strange concludes that the heterogeneity of the results make it difficult to draw firm conclusions. She does claim that identification training was the only type to show significant change in discrimination performance, and rejects Lane's claim that speech perception can be quickly and easily modified in the laboratory.

The claim that speech perception is categorical and that therefore "subjects can only discriminate between stimuli that they can identify differently" (Studdert-Kennedy, Liberman, Harris & Cooper, 1970) has also inspired training studies. Carney et al (1977) trained three subjects on a /ba/-/pa/ VOT continuum using a discrimination task with immediate feedback. After 12-13 training sessions of 1.5 hours each, the experimenters found that the subjects had improved in their ability to discriminate between same-category stimuli. In a separate experiment, in which subjects performed both a discrimination and identification task on the same stimuli, the authors find that the mixed task has no significant influence on discrimination or identification performance, and so conclude that the two tasks do not operate in different modes. However, in addition to the small number of subjects, these subjects have elsewhere been described as "veterans of psychophysical research" (McClaskey, Pisoni & Carrell, 1983). A similar study (Samuel, 1977) used an adaptive training technique in which the difference in VOT between two stimuli used in a discrimination task was modified based on subject performance. As the subject improved, the VOT difference was lessened. Samuel notes that even before training subjects had some ability to discriminate between same-category stimuli, and that training for the three subjects (which include himself) led to significant improvements in that ability. Samuel argues that these results favor the idea that both an acoustic and phonetic mode are active in speech perception, as per Pisoni (1973). Edman (1980) also used training in an effort to improve the intra-phonemic discrimination abilities of his subjects. In three different training experiments using a fixed standard AX discrimination task (the A stimulus is always the same, while the X stimulus varies) with immediate feedback, he found significant improvements using both a VOT series and a place of articulation contrast representing a bilabial-alveolar-velar continuum (both voiced and unvoiced versions). Edman used lengthy training periods that concluded only when subjects had reached a pre-defined criterion performance. Subjects received an average of about 10 hours of training in one-hour sessions over two weeks. The intensive nature of the training most likely explains the robust improvements he found (though Edman also had only a small number of subjects, 6 or less, and found substantial inter-subject variability). Intriguingly, Edman was also concerned with what was being learned and thus included tests of transfer in each of his training studies. He found that discrimination performance transferred from a bilabial VOT series to a velar VOT series and vice versa, from a voiced bilabial-alveolar-velar place of articulation series to an unvoiced series, and from one set of synthetic bilabial VOT stimuli to a very different set. Based on these results, Edman concludes that the learning displayed by subjects must have been, to at least some degree, phonetic (rather than wholly stimulus specific) or transfer could not have occurred. However, since transfer was not perfect (subjects showed greater improvements in the stimuli they were trained on) the learning "can not be regarded as purely phonetic." (Edman, 1980:70)

The relationship between these two modes of processing is also addressed by Repp (1981) who attempted to train subjects to "relinquish the phonetic mode of processing" (Repp, 1981:222). In one experiment, 10 subjects were trained with a discrimination task using fricative noise that is perceived as /s/ or /ʃ/ with a following vowel, but is perceived as nonlinguistic when isolated. Although the subjects developed noncategorical perception of the stimuli with training, they reverted to categorical when tested with the sounds made whole again (frication plus vowel). A second experiment trained 7 subjects (3 from the earlier experiment) using the frication without vowel and frication with vowel together hoping that it would help the subjects realize the differences that existed in the full syllable stimuli, and found a fair degree of variability in subjects, some of whom learned to perceive the acoustic differences, even in the linguistic context. Repp concludes that the normal case is for speech to be perceived in a special phonetic mode, but that explicit instructions and tasks could allow some individuals to utilize the auditory mode in processing speech sounds.

Pisoni, Aslin, Perey & Hennesy (1982) used training on native speakers of English to determine if they could be trained to perceive 3 categories of voicing, as are found in the Thai language. In English, initial bilabials are either pre-voiced (like /b/) or post-voiced such that a brief period of aspiration is heard after the release of the labial closure in /p/. Thus, when the stop is released simultaneously to the beginning of voicing, NS of English will generally perceive this as a pre-voiced /b/. However, in Thai this represents a separate phoneme. When /p/ comes in a consonant cluster, such as in a word like /spɪt/, it is not aspirated so pre-voiced and voiced stops represent allophonic variations of one another in English. Pisoni et al (1982) wanted to determine if NS of English could learn to segment the VOT continuum as NS of Thai presumably would. They gave 12 subjects identification training for three different stimuli on the voicing continuum (-70, 0 and +70) and gave the subjects immediate feedback. After several hours of training, the six best learners were invited back for further training the next day. After further training, these subjects labeled three distinct categories and showed discrimination peaks at both their normal English boundary, and at the new Thai boundary. They conclude "the perceptual mechanisms used by adults in categorizing stop consonants can be modified easily with simple laboratory techniques in a short amount of time" (Pisoni et al, 1982:297). McClaskey et al (1983) repeated the experiment above, but also tested generalization of training from a bilabial to alveolar series and vice versa. Like Edman (1980), they found that performance on the untrained series did improve, suggesting that the learning is not merely stimulus specific.

Concurrent with these laboratory training experiments, studies in cross language speech perception, using real speech tokens as stimuli, revealed that second language learners had great difficulty learning to perceive some foreign language phonemes, even after long experience with the language. Borrowing from the training techniques developed in the psychophysical experiments described above, Strange & Dittman (1984) conducted a training experiment designed to see if NS of Japanese could improve their perception of real speech tokens featuring the English /r/-/l/ contrast. Because of its importance to the present pedagogical research, and because it set the stage for much of the training research to follow, this experiment will be discussed in detail here.

Strange & Dittman begin by noting that nearly all prior training studies had concerned themselves exclusively with the VOT dimension, an acoustic cue familiar to nearly all users of language. The case of training NS of Japanese to perceive the /r/-/l/ contrast however, requires learning to discriminate between two phones, neither of which is found in the native language. Furthermore, the /r/-/l/ discrimination requires attention to a complex signal with varying temporal and spectral parameters. Finally, cross-language speech perception studies with real speech examples had shown that difficulties with the discrimination persist for many Japanese learners, even after long immersion in the English language (Goto, 1971; MacKain, 1981; Miyawaki, 1995). Thus, training Japanese learners to perceive the /r/-l/ contrast is expected to be more difficult than altering perception of a simple VOT distinction. Strange & Dittman (1984) generated a synthetic continuum of words from rock to lock to use in training. The pre- and posttests included identification and discrimination tests with these stimuli, a second synthetic stimuli of rake-lake and an identification test with real speech minimal pairs featuring the contrast in multiple phonetic contexts. Subjects (there were 8 subjects, four of whom first served as the control group and then took training) performed an AX discrimination task and received immediate feedback. Each subject received between 14-18 training sessions and each session consisted of 7 blocks of 18 trials such that subjects performed between 1764 and 2268 total discriminations. Seven of eight subjects improved significantly on the trained stimuli at posttest and showed evidence that their perception of the contrast had become more categorical (i.e. more native like). Of those seven, five also showed significant improvement in identifying stimuli in the rake/lake series as well. Crucially however, there was no evidence for improvement in identifying /r/ and /l/ as found in real speech tokens. This failure to transfer is of crucial importance to a pedagogically motivated phonetic training program. If the gains of training do not extend beyond the training stimuli themselves, then the training is of no value. Strange & Dittman note that given the parameters of their training, it is unsurprising that training gains did not generalize to real speech tokens. They conclude that:

> future studies should be designed in such a way that subjects learn to abstract the relevant parameters which differentiate the phonemes while ignoring the acoustic and contextual variations that are not distinctive with respect to the contrast. This would include training on the contrast with more than one set of stimuli and in more than one phonetic context. (141)

Thus, although Strange & Dittman's study failed to show significant improvement in the identification of /r/ and /l/ in real speech, it laid the groundwork for the successful training studies that would follow.

In a series of experiments, Jamieson & Morosan (1986; 1989; Morosan & Jamieson, 1989) trained Canadian Francophones to discriminate between English /ð/ and /θ/. They also used synthetic stimuli for training, but made two modifications to the training procedure. First, they had the subjects perform an identification task. They reasoned that practicing at discrimination "is likely to have the undesired effect of enhancing sensitivity to within-category acoustic differences." (Jamieson & Morosan, 1986:207). They also used an adaptive technique in which early training stimuli had longer periods of frication, and therefore were easier to identify, that became progressively shorter as subjects improved. Training time was dependent on subject performance, but in no case did it exceed 90 minutes. The experimenters found significant improvement in identifying real speech stimuli after the training period in comparison with a control group. In a follow up experiment (Jamieson & Morosan, 1989) they used prototype training (without a gradual increase in task difficulty) and found it to be inferior, though they do note that the subjects that received adaptive training started at a much lower initial level, and so had more room for improvement, and that there were no significant differences in the two techniques in improving identification of real speech stimuli. A final experiment (Morosan & Jamieson, 1989) found that identification improvements in CV (consonant + vowel) contexts failed to transfer to VCV or VC contexts and that discrimination between /ð/ and /d/ did not improve as a result of training on the /ð/-/θ/ distinction. Thus, although the training did generalize to real speech tokens, it did not generalize to other allophonic variants.

The results of Strange & Dittman (1984) were further built upon by a series of experiments that established that phonetic training could be used to improve Japanese perception of the /r/-/l/ contrast such that the trainees new-found skills generalized to real speech tokens that were not part of the training set (Logan, et al, 1991; Lively, Logan, & Pisoni, 1993; Lively, Pisoni, Yamada, Tohkura, & Yamada, 1994). The results of these experiments have provided the methodological foundation for nearly all subsequent phonetic training experiments, so it is once again worth reviewing the work in detail.

The first of the experiments (Logan et al, 1991) made three important modifications to the Strange & Dittman (1984) procedure. First they used an identification task for training, reasoning that this would encourage subjects to classify the stimuli into categories. Second, the training stimuli featured the /r/-/l/ contrast in multiple phonetic contexts, rather than word initial only, as in the Strange & Dittman (1984) experiment. This relates to the concept of allophonic variation, discussed previously, in which different instances on the same phoneme may have different articulatory and/or acoustic correlates due to phonetic context. Therefore, the critical cues for discriminating /r/ from /l/ may be different in different contexts and so learning to make the discrimination in one context will not necessarily transfer to others. Finally, the training stimuli were real speech tokens supplied by five different native speakers of English. This decision relates to the concept of normalization in speech perception discussed earlier. The experimenters reason that, in order to properly engage the normalization process in the subjects, it was necessary not only to expose them to the acoustic differences relevant to the contrast, but also to acoustic differences specific to individual speakers that are irrelevant to the contrast. With both sources of information, subjects could form phonetic categories that are sufficiently robust to generalize to a completely new voice that is not part of the training set.

The subjects for Logan et al (1991) were 6 NS of Japanese who had been living in the United States from 3 months to 6 years. Subjects participated in 15 training sessions over three weeks and each session included 272 identification trials for a total of 4080 identifications. Immediate feedback was given and the stimulus was repeated for any incorrect answer. During any one session, only one training voice was heard. After training, the subjects showed a significant, though small, increase in their ability to identify /r/ or /l/ in stimuli that were not part of the training set. This methodology became known as High Variability Phonetic Training (HVPT) and, as shall be seen, has been used to successfully train language learners from a variety of different L1s to perceive phonetic contrasts found in a variety of L2s. With some minor modifications, this is the methodology that we will employ in training our Arabic learners as described in Chapter 5.

Lively et al (1993) used the same training methodology above; however, they removed stimuli that featured the /r/-/l/ contrast in final and final cluster position as they found that many subjects were already near ceiling for these contexts. They also had one training group train on a single talker in an effort to confirm the importance of stimulus variability. Interestingly, not only did the subjects who heard only one training voice fail to generalize their new perceptive abilities to new voices, but even to new tokens from the same voice. Thus, their new perceptual abilities were specific to the stimuli used in training and it could not be claimed that any actual speech learning had occurred. They conclude that without sufficient stimulus diversity, learners cannot form robust categories for foreign language phonemes.

The final experiment in the series (Lively et al, 1994) trained Japanese who were still living in Japan and had minimal exposure to spoken English. They found that the training was still effective with these trainees and that the positive effects of training lasted for at least 6 months, even in the absence of no further exposure to English, lending further evidence that the training procedure was allowing trainees to form robust categories for the English language phonemes.

Taken together the results of the training studies described above demonstrate the pedagogical potential of HVPT. They show that the learning of difficult phonetic distinctions can be quickly improved with relatively simple laboratory techniques. They show that these improvements are true speech learning in that they generalize to speech tokens that were not part of the training set (though whether or not the newfound perceptual abilities can be engaged in actual foreign language listening remains an open question). Finally, they show that these improvements do not disappear with the end of training, but can persist for some time. It therefore seems warranted to explore the possibility of harnessing these methods for pedagogical goals. The remainder of this chapter will review the important findings of training studies that have been conducted since the pioneering experiments described above to give a more complete understanding of the best practices, promise and limitations of phonetic training and also to see how such experiments can, in addition to their pedagogical value, shed light on important questions of speech perception and second language acquisition.

4.2 Best practices in phonetic training

In their 1978 review article of phonetic training research that had been conducted up to that time Strange & Jenkins conclude that "laboratory training studies have had little effect on listeners' ability to discriminate distinctions not employed in the listeners' native language" (162). Thirty years later, an article with the same purpose maintains that "Laboratory training can lead to successful non-native contrast learning even for the most difficult cases"(300: Bradlow, 2008). In the intervening period a number of phonetic training studies with a variety of different L1/L2 combinations have been published that confirm and build upon the results of the studies described above. In what follows in this chapter, these studies are examined with the aim of:

1. Describing the components of a methodologically sound, effective and maximally informative phonetic training study that can reliably contribute to our knowledge of cross language speech perception,

2. Surveying the types of second language speech contrasts that have been the subject of these studies and drawing conclusions that can be made about the ease or difficulty of inducing perceptual change in different types speech contrasts such as consonants and vowels, quality and quantity contrasts and suprasegmental contrasts such a Chinese tones. This section will also address the potential influence exerted by the native language on speech language learning, which we have seen is critical in considering the pattern of findings in L2 speech learning.
3. Considering the implications of these findings for the models of L2 speech perception that we have previously discussed such as the Speech Learning Model (SLM), the Perceptual Assimilation Model (PAM & PAM-L2), the Native Language Magnet (NLM) theory as well as ideas about L2 speech learning addressed more specifically to uncovering the mechanisms that drive L2 speech learning.

We do not, of course, probe each and every one of these many issues in the empirical study described in Chapter 5; nevertheless, we believe that a complete survey of the literature is of value for a number of reasons. For one, proper design of the current study can benefit greatly from a thorough understanding of the studies that precede it, thus, even though we will not be comparing, for example, the differential effects of training on Arabic phonemes on learners of two different native language backgrounds, it is still of value to consider studies that do so, as the information generated by them may be useful in guiding our own predictions. Two, as shall become clear in the following discussion, there are surprisingly few studies that meet all of the criteria that we shall identify as necessary for a maximally informative study. Only by a thorough survey of prior studies do those criteria become clear and thereby aid us in both evaluating the reported results of other studies, but also in designing a methodologically sound study whose results we can be confident in defending before a potentially skeptical audience. Third, the Arabic phonetic training study described here has never been intended to be the final word on the topic. Rather, the goal is to demonstrate empirically that such training can work for Arabic and that it can do so in a way that makes it an efficient and effective pedagogical tool to be used alongside a communicative, proficiency-based Arabic curriculum for adult native speakers of English learning this language. By taking stock of all the phonetic training studies that have preceded this one, we can keep an eye to the future as to what potential modifications and additions may be made to increase the effectiveness of the training and to expand it to include other perceptual problems that are known to vex learners of Arabic. Finally, although an excellent recent review of phonetic training does exist (Bradlow, 2008), it restricts itself exclusively to consideration of training of a single contrast (English /r/-/l/) and examines learning in subjects from a single native speaker background (Japanese). A more complete understanding of phonetic training can be had through consideration of the many other types of contrasts and native language backgrounds that have been studied.

4.2.1 Important elements of study design

4.2.1.1 Measuring the effectiveness of training

As indicated, both proper design of a future phonetic training study and proper evaluation of prior phonetic training studies require a consideration of those factors that contribute to a maximally informative study. Whether the goal is strictly pedagogical, or designed to evaluate theories of second language speech learning, phonetic training studies ask if listeners' perception of speech can be altered in the short-term using relatively simple perceptual tasks without the explicit provision of meta-linguistic information. Thus a typical research question might be phrased as follows "Can a HVPT training program improve the ability of native speakers of Spanish to accurately identify the English vowels /i/ and /ɪ/?" For a PT study to provide evidence that it can, it must minimally show significant differences in listener performance on a given perceptual task between a trained group and a control group of untrained learners. As we have seen in the Strange & Dittman (1984) study and experiment 2 from Lively et al (1993) it is not a foregone conclusion that improvement in the training task, with the training stimuli, constitutes speech learning. In both cases the learning demonstrated by the subjects proved to be stimulus specific. Subjects failed to show that they had learned the critical acoustic cues that differentiate the English /r/-/l/ contrast in such a way that might allow them to apply their new abilities performing real world language tasks. Recall from our discussion of speaker normalization in Chapter 2 that native speakers of a given language are able to rapidly make adjustments in their perception of speech sounds according to information they have about the individual characteristics of a given speakers voice. Among other things, this allows them to recover the same message spoken in a very high voice compared to a very low voice despite the vast acoustic differences between the two. In the absence of this ability, linguistic communication would be impossible. Therefore, in order to make the claim that what trainees have learned is linguistic in nature, we must be able to show that they are capable of speaker normalization in perceiving the new contrast that they have learned. The best way to show this is through the use of a test for generalization to a new speaker as first recommended by Strange & Dittman (1984), and such a test should be considered the primary measure of the success of a given training program in altering second language speech. Given that the extra effort required for such a test is minimal measured relative to the total time invested in a PT study by researchers and subjects alike, such a test should be considered an obligatory component of any such study.

Unfortunately, even since the Strange & Dittman (1984) study, many PT studies have failed to include this all-important feature. This appears to be especially true of studies that explore explicit mechanisms of L2 speech learning (Maye & Gerken, 2000; Francis, Baldwin & Nusbaum, 2000; Maye & Gerken, 2001; Francis & Nusbaum, 2002; McCandliss et al, 2001; Hayes-Harb, 2007). As a consequence, the results of these studies, although certainly intriguing, require verification before their conclusions can be applied to our knowledge of L2 speech learning. Despite this drawback, the results of these studies will be further explored in later sections of this chapter with the caveat that any conclusions drawn based upon them must be considered preliminary. A number of other studies include the test of generalization only at posttest (Wang, Spence, Jongman & Sereno, 1999; Wang & Munro, 2004; Lengeris, 2008; Motohashi-Saigo & Hardison, 2009), presumably so that it will be the first contact that the subjects have with the stimuli and the subjects' results will not be influenced by their prior exposure. This, however, does not seem appropriate. For one, several studies have shown differential subject performance on different native speaker voices (e.g. Logan et al, 1991; Takagi, 2002); so including the test of generalization at pretest allows measurement of the size of the training effect. Furthermore, since it seems that exposure to prototype (i.e. unaltered natural speech tokens) stimuli in the absence of feedback does not lead to perceptual learning (McCandliss et al, 2002) the inclusion of the generalization test prior to training is unlikely to drastically bias the results at posttest.

Although it is the most important, the test of generalization to a new voice is not the only meaningful measure of the robustness of speech learning from phonetic training. Pruitt, Jenkins & Strange (2006) identify six additional tests of generalization (they refer to it as transfer) that “differ in the amount and type of nonrelevant variation that the listener must ignore to make a correct response.” (1689) They include 1) Generalization to a new task (from identification to discrimination), 2) Generalization to new stimuli from the same speaker, 3) Generalization to a new phonetic context (e.g. consonant + /i/ to consonant + /u/, 4) Generalization to new syllabic environments (e.g. from word initial to word final), 5) Generalization to new contrasts with the same feature (e.g. identification of voicing contrast from bilabial to alveolar place of articulation) and 6) Generalization from isolated tokens to tokens presented in a more realistic context (i.e. running speech). Of course, not every PT study can be expected to test each of these types of generalization. Each individual study should consider the constraints and goals of their study when deciding which tests to include. Other things being equal, the more types of generalization are tested, the more informative the results will be.

Evidence of robust learning can also be provided by tests of retention, given several months after training has ended. Such tests can help determine if meaningful and long term perceptual learning have occurred, rather than a temporary alteration of behavioral responses based on immediate task demands. In general, tests of retention have shown that the learning induced by HVTP does remain at tests of retention (Lively et al, 1994 – 3 & 6 months; Kraus, McGee, Carrell, King, Tremblay & Nicol, 1995 – 1 month; Wang et al, 1999 – 6 months; Bradlow, Akahane-Yamada, Pisoni & Tohkura, 1999 – 3 months; Wang & Munro, 2004 – 3 months, Iverson & Evans, 2009 – 4 ½ months).

Both the SLM and the PAM-L2 posit a correlation between L2 speech perception and L2 speech production, so a further test of the robustness of learning can be provided by measures of pre- and posttest productions of the speech contrasts under study. Many

PT studies have incorporated such tests, and these will be reviewed in more detail in the final section of this chapter.

4.2.1.2 Control groups

A second indispensable component of a methodologically sound PT study is the inclusion of a control group, a subset of subjects that takes the pretest and posttest without training, with the interest interval equivalent to the training period. Despite the fundamental nature of the use of control groups in scientific study, this component too is often missing from PT studies (for example Maye & Gerken, 2000; Francis, Baldwin & Nusbaum, 2000; Wayland & Guion, 2004; Hazan et al, 2005; Kelly, Hirata, Simester, Burch, Cullings & Demakakos, 2008). There are two main reasons that use of an untrained control group is necessary. First, it cannot be assumed that no learning will occur in the absence of training. This is particularly true when subjects are living in the L2 environment or currently studying it, and even more so when they are beginning learners of the language when learning of new contrasts is most likely to be most rapid. Even in the event that subjects presumed to be at asymptote are used, it is necessary to use an untrained control group in order to control for the potential effects of task familiarity. Many PT studies that do use untrained control groups show statistically significant improvements from pretest to posttest for untrained subjects, indicating either that the task itself induces some perceptual learning, or that the task becomes easier with familiarity. In one study (Hayes-Harb, 2007), to be reviewed in more detail later, the untrained control group outperformed one of the training groups, suggesting that the training methodology had suppressed previously existing discrimination ability. In the absence of a control group, this rather interesting result might have been missed. Assignment into control and training groups should be pseudorandom such that the measures of the central tendency of the groups' pretest performance are roughly equal. After training, the efficacy of the training method is assessed by a comparison of improvement in the trained group with improvement in the control group.

4.2.1.3 Comparisons with native speakers

It is also very useful to include tests of native speakers on the pre- and posttest stimuli, or at the very least to have native speakers listen to the stimuli and insure that they are in fact discriminable. It should not be assumed that native speakers will perform at 100% on all test items, and it may be that there are some contrasts, in some phonetic contexts for which NS performance deviates significantly from 100%. If the perception test is sufficiently difficult, such that the distribution of NS scores is normally distributed, it is possible to determine if any individual subjects demonstrate native-like performance either before or after training.

4.2.1.4 Training task

The test of generalization and the use of an untrained control group are the two most critical components of a PT study investigating L2 speech learning. In the remainder of this section, we discuss other aspects of a PT study and how they potentially impact the efficacy of the training and the conclusions that can be drawn from it. One of the most important considerations in designing a study is the task that subjects will perform as part of the training. In the first section of this chapter we noted that there are two main task types that have been used in PT: discrimination tasks and identification tasks. In addition to ABX discrimination, described in section 1, researchers have used AX discrimination, oddball paradigms and identification tasks for testing and training of L2 perception. Identification tasks differ in the number of potential categories to which a given stimulus may be assigned. It is usually the case, especially with consonant or quantity training, that only two choices are provided. With vowels it is more common to see a larger set of possible categories. In that event it may be important to consider the range of possible confusions, a fact demonstrated by a pair of experiments by Nishi & Kewley-Port (2007; 2008). In their studies, the authors trained both Japanese (2007) and Korean (2008) native speakers in the perception of 9 American English vowels (/iː/-/ɪ/-/ɛ/-/æː-/ɑː/-/ʌ/-/ɔː/-/ʊ/-/uː/) and were particularly interested in how the variation in stimulus sets influenced the learning process. In the first study, with Japanese learners (Nishi & Kewley-Port, 2007), their trained subjects were divided into two groups, one of which received training on all 9 vowels and the training task was a 9-choice identification task, while the second group was trained only on the those vowels found to be most difficult in a pilot study (/ɑː/-/ʌ/-/ʊ/). The training was based on the HVPT methodology, though only two different voices, rather than the more typical five, were used in training. The most significant finding of the study is that the restricted set group only improved in identifying the three vowels that were part of their training. An even more important result was found in the second study with Korean learners (Nishi & Kewley-Port, 2008).

In this experiment there were three training groups. Subjects in the 9-9 condition were exposed to training sets with the full set of nine vowels for each of two phases of training. The training sets for subjects in the 3-9 condition consisted of only three vowels for the first phase, and all nine for the second. Finally, in the 9-3 condition the first phase of training featured all nine vowels, but the second phase only three. They found that the subjects in the 3-9 condition showed rapid progress in identifying /ʊ/ during the first phase of the training, but when switched to the full set training they lost that ability and failed to recover it. The same was not true of the other two groups and the 9-9 group showed the most improvement overall. This result suggests that subjects in the 3-9 condition had learned to use a cue for identifying /ʊ/ in the first part of training that was not sufficient for uniquely identifying it when compared to the full set of vowels. Moreover, having latched onto this cue, they proved unable to redirect their attention to the acoustic cues that were in fact necessary for identifying /ʊ/. The authors note that confirmation of this hypothesis is necessary, but if it holds it suggests that the identification task used in training identification of a given phonetic feature should include all possible competing phonemes so that subjects learn the appropriate cues for the identification.

It has frequently been asserted that discrimination tasks are inappropriate for phonetic training since they focus the subjects attention on with-category variation, rather than on between category variation, which is necessary for the formation of new phonetic categories (Logan et al, 1991; Jamieson & Morosan, 1986; Logan & Pruitt, 1995). Consequently, most studies have used an identification task in training. It appears to be the case that only one study (Flege, 1995b) has actually put that claim to the test, by comparing the two types of training side by side while holding other factors constant. Rather than an identity discrimination task, in which the subject must decide if the two stimuli are physically identical, Flege employed a category discrimination task, in which subjects decide if the two stimuli came from the same category. In this case, two different speakers provide the two stimuli being compared. This task avoids the problem of focusing subject attention on within stimulus differences and forces them to focus on the cues that determine category membership while ignoring other differences in the stimuli. Flege trained native speakers of Mandarin to perceive word final voicing in English words (voicing contrasts exist in Mandarin, but phonotactic constraints convert all voiced phonemes to unvoiced at the end of words), and found that both groups showed significant improvement at posttest compared to an untrained control group, and that there were no significant differences between the two. In fact, in the test of retention there was a slight advantage for the discrimination-trained group. However, given that VOT perception has been shown to be relatively easy to manipulate (McClaskey et al 1983), and that Mandarin speakers are already quite familiar with the distinction in other phonetic contexts, this may not be a particularly strong test of the claim that identification training is better. Furthermore, subjects from the identification-trained group indicated more confidence in the gains they made and a greater willingness to continue training than the discrimination trained group, a finding of considerable relevance to pedagogical implementation of PT.

Some credence to the idea that identification training leads to more native-like phoneme representations is provided by Guenther, Husain, Cohen & Shinn-Cunningham (1999). This study compared discrimination training with identification training in learning to discriminate between nonspeech sounds (two different 'categories' of nonspeech white noise with different center frequencies) and, as predicted, identification training led to a significant decrease in subjects within category discrimination abilities, as is found in the categorical perception of phonemes, while discrimination training led to an increase in subject sensitivity to within-category variation, as found in the perception of nonspeech sounds. Of course there was no test of generalization to a new voice, so it is uncertain if the same results would obtain with speech stimuli.

Neither of two studies that use oddball category discrimination in training, one that trained NS of German to perceive the length contrasts of Japanese (Menning, Imaizumi, Zwisterlood & Pantev, 2002), and another that trained Thai tone contrasts to NS of Mandarin and English (Wayland & Guion, 2004), include tests of generalization to a new voice, so conclusions that can be drawn from them are limited. It is noteworthy though that the Wayland & Guion (2004) study failed to induce significant training gains in the NS of English subjects despite using 1200 total training trials (though they did in the Mandarin speaking subjects), while another study, using identification training and only 720 total trials, was able to successfully improve NS of English perception of Mandarin tones (Wang, Spence, Jongman & Sereno, 1999). Again, no firm conclusions can be drawn from these results and despite assertions to the contrary, the question of the effectiveness of category discrimination training remains very much an open one.

4.2.1.5 Training stimuli

In addition to the type of task that will be used in training, researchers must consider the nature of the stimuli that will be used. Most commonly, researchers have followed the lead of Logan et al (1991) and used real-speech tokens provided by NS of the language from which the trained contrast is taken. As discussed in the description of this study, the use of real speech tokens from multiple talkers is considered critical for inducing robust perceptual learning that transfers to other voices. This is because it gives trainees the opportunity to determine which patterns of variation in the speech signal are critical to the distinction, as this should be common to all voices, and which patterns are unique to specific speakers and should be ignored. Thus, it seems safe to conclude that training stimuli should at least begin as real speech tokens. We do not discount the possibility that as speech synthesis technologies and techniques continue to improve, and as our knowledge of how subjects from different L1s weight the importance of different speech cues and how those weights may be altered through experience, it may be possible to synthesize training stimuli that are superior to natural speech in inducing changes in trainees perception, however, given the labor-intensive nature of speech synthesis, we have not reached that time yet.

Although training stimuli should not be wholly synthetic, some researchers have had success using resynthesized real speech tokens such that certain acoustic features are made more salient. This technique, which has been described as both ‘fading’ and ‘adaptive’ (we shall use the latter), was first used by Jamieson & Morosan (1986) in training French speakers to discriminate between English /θ/ and /ð/. The idea is to alter the speech signal so that acoustic cues relevant to the contrast are made more salient. This draws the subjects’ attention to these cues and as they progress through training the amount of exaggeration is slowly reduced until the tokens are representative of natural speech. In the initial Jamieson & Morosan (1986) study, the exaggeration was a simple lengthening of the fricative portion of the stimuli but subsequent researchers have used more sophisticated techniques to change the quality of the acoustic signal (Protopapas & Calhoun, 2000; McCandliss et al 2002; Wang & Munro, 2004; Iverson, Hazan & Bannister, 2005). Unfortunately, few studies that use adaptive training compare its efficacy to prototype training, which uses unaltered speech tokens. Those that do have not found a substantial advantage for adaptive training. Both Jamieson & Morosan (1989) and Morosan & Jamieson (1989) train Canadian francophones on the perception of English /θ/ and /ð/, the former with prototype training, the latter with adaptive. They find an advantage for adaptive training in identifying synthetic tokens, but the two groups were not significantly different in identifying real speech tokens. McCandliss et al (2002), who do not include a test of generalization, trained NS of Japanese on the /r/-/l/ contrast of English and demonstrate perceptual gains in the adaptive group, even without feedback. However, they did not outperform a group trained with feedback in the prototype condition. Iverson et al (2005) compare prototype training with a variety of different cue manipulations and find no advantage for cue manipulation so conclude that the prototype method is superior because it is considerably less labor intensive.

We have previously discussed (see Chapter 2) the difference between phonemes and phonetic categories and the notion of allophonic variation, and these concepts must be taken into account when selecting stimuli for a phonetic training program. Although PT studies have shown examples of certain phonological features, most notably VOT, transferring from a trained place of articulation to an untrained one (Rochet, 1995; Pruitt, Jenkins & Strange, 2006), PT studies have generally failed to find transfer of training gains from one word position or vocalic context to another (Morosan & Jamieson 1989; Lively et al, 1993), or have found transfer that is significantly weaker than the gains found for the trained contexts (Iverson, Hazan & Bannister, 2005; Pruitt et al, 2006). This being the case, the design of a PT study must take care to include stimuli from all of the phonetic contexts for which training gains are sought. It should also be noted though that with NS of Japanese perception of the /r/-/l/ contrast of English, pretest measures have found that certain phonetic contexts are much more difficult than others and that training on the contrast in the word final position is unnecessary (Lively et al, 1993). While NS of Korean also have difficulty with the /r/-/l/ contrast, they have been found to have the most difficulty with word initial contexts (Ingram & Park, 1998). While the contrast types that are most difficult for learners of a given language background may be determined based on anecdotal evidence from teachers and the learners themselves, uncovering the specific phonetic contexts that are most difficult is likely to require empirical testing. The use of pilot tests on learner perception can be helpful in improving the efficiency of a PT study. In the event that such information is not available, or when there is substantial substantial what? within subject variation on the contrasts that are most difficult, a training regime that alters the proportion of stimuli from different phonetic contexts in response to subject performance may be an alternative solution.

A final component that researches may wish to consider when preparing stimuli is the incorporation of the visual modality, that is, allowing learners to see, as well as hear, the stimuli as they are spoken. This follows from the close linkage between perception and production of phonemes in L2 learners that was discussed in Chapter 3. Since studies incorporating the visual modality will be discussed below, when the implications of PT research on theories of L2 speech learning are considered, they will not be described here.

4.2.1.5 Feedback

The use of feedback must also be considered when designing a PT study. Although McCandliss et al (2002) showed successful training in the absence of feedback when subjects underwent adaptive training, the group that received prototype training, but no feedback, failed to learn. Furthermore, since feedback does not seem to hinder learning in any circumstances, and since subjects are likely to prefer to get feedback, there is no reason to leave it out. The question then is how feedback should be provided. Most studies simply indicate to subjects whether or not their answer was correct, and frequently repeat the stimulus when a wrong answer is given. A review article of PT studies (Logan & Pruitt, 1995) suggests that, in addition to indicating to subjects whether or not they were correct, it may be of value to provide subjects with reaction time information as well, allowing them to consider the potential tradeoffs between speed and accuracy. This possibility remains untested and it is also uncertain if such a tradeoff exists. Since no studies have compared the relative efficacy of different types and amounts of feedback, it is difficult to draw conclusions about what may and may not work. From a pedagogical perspective, it makes sense to give subjects as much control over the feedback they receive as possible. Allowing them, for example, to rehear the trial stimulus, its counterpart, and perhaps even other versions of the same contrast so that they could spend time on the contrasts that were most difficult to them. Logan & Pruitt (1995) note that interval feedback, such as indication to subjects their per cent correct score after each training session, may increase learner motivation, also of value in a pedagogical implementation of PT. In a similar vein, it may be valuable to present learners with the option of viewing a progress chart that tracks and displays session by session improvement (or lack thereof).

4.2.1.6 Training time

Of obvious interest to the researcher, especially when pedagogy is the goal, is how much training is necessary. The answer to that question is likely to vary according to the contrast under study, the L1 of the trainees and factors unique to individual subjects. Mapping the time course of training can also be difficult because although it is easy to track the improvement of subjects' performance on the training stimuli, the more important test of generalization can only be given at limited intervals so that subjects do not become too familiar with what is supposed to be a novel voice. The training of NS of Japanese on the English /r/-/l/ contrast is the only case where enough studies have been published that we can draw some tentative conclusions about appropriate training time. Trainees in Lively et al (1993) received a total of 4080 identification trials spread over three weeks of training. The experimenters noted that while accuracy scores seemed to peak after two weeks of training, reaction times continued to drop in the third week. Yamada (1993) used the same stimuli and procedures as Logan et al (1991) and Lively et al (1993), but tripled the amount of training. She records substantially more improvement in the trained group as a whole compared to these earlier studies. She also reports that while some subjects plateaued before the training ended, other continued to improve throughout training. Protopapas & Calhoun (2000), who used adaptive training, report more substantial improvement than the Logan et al (1991) study in fewer trials (750-1000), though this study considers only word initial /r/-/l/ and fails to include an untrained control group. Takagi (2000), who focuses on individual variation in response to training, finds large individual differences in the time course of training. Sinnott, Gonzales, Masood & Ishiara (2007) used a psychoactive training procedure that had previously been used successfully in training monkeys to perceive phonemic contrasts. This involves repeated presentations of the stimulus until the subject makes a response and the repetition of individual trials after an incorrect response. Training, using 4 voices, lasted 8 weeks and the researchers report that subjects reached native-like levels, though they do not give a test of generalization to a non-training voice. Studies investigating the effects of the statistical distribution of speech sounds in the environment (Maye & Gerken, 2000; Maye & Gerken, 2001; Hayes-Harb. 2007) report modification of VOT perception in mere minutes of exposure, and generalization to a new place of articulation (Maye & Gerken, 2001), though they also fail to test for generalization to new voices.

Given that all PT studies have found substantial within-subject variation, both at pretest, and in response to training, at present, the only recommendation that can be made regarding the appropriate number of trials is that individual performance be tracked, and training stopped when a plateau appears to have been reached. As we will see though from individual training data in Chapter 5, identifying a plateau is unlikely to be easy and may vary considerably from one subject to another.

4.2.2 The scope of phonetic training

Since the Strange & Dittman (1984) article a large number of studies have been published claiming successful acquisition of a given L2 speech feature by a given group of L1 learners, leaving the impression that training never fails. This, however, may also be due to the bias of the scientific publishing process, which usually refuses to publish negative results. It is not impossible to imagine that a large number of unsuccessful training studies, though still valuable research, remain unknown to us. This is an unfortunate oversight, and is hardly unique to the field of phonetic training. For the time being, we can only take advantage of those studies that have been published, but in doing so should not lose sight of the fact that we may not have a complete picture of the research that has been conducted.

Generally speaking, phonetic training studies have addressed themselves to one of four different features – consonants, vowels, quantity (or length) and suprasegmentals (tones). The differences in how they have been approached, and the conclusions that have been drawn, are not profoundly different. Since the training study that we present in Chapter 5 attempts to improve phonemic perception of consonants, we will restrict our discussion to training studies that focus on consonants.

As we have indicated, by far the most studied case of L2 speech learning is the Japanese acquisition of the English /r/-/l/ contrast, which was first chosen by Strange & Dittman (1984) because substantial literature already existed describing the problem, and because they saw it as a more robust test of perceptual learning than the VOT contrasts that had been studied up to that point. Because neither /r/ nor /l/ were found in Japanese phonology, and because of the complex phonological, phonetic and articulatory differences between /r/ and /l/, they saw it as a litmus test of the potential for phonetic training to alter L2 learning outcomes. In her summary of PT of this contrast to NS of Japanese, Bradlow (2008) writes:

> the English /r/–/l/ contrast presents great difficulty for native Japanese speakers due to extensive mismatches between the underlying systems of contrasting approximant categories of the two languages, the particular acoustic-phonetic features that listeners of the two languages have learned to attend to, and the articulatory configurations that talkers of the two languages have learned to produce. (293)

Nevertheless, as previously described, a large number of studies have shown successful training of this contrast to NS of Japan in tests of generalization to a new voice (Logan et al, 1991; Lively et al, 1993; Yamada, 1993; Bradlow et al, 1997; Bradlow et al, 1999; Protopapas & Calhoun, 2000; Takagi, 2002; Hardison, 2003; Hazan et al, 2005; Iverson et al, 2005).

Efforts have also been made to train NS of English to perceive the Hindi dental/retroflex contrast. Retroflex articulations involve a curling back of the tongue to contact the front portion of the palate and in Hindi the retroflex /ʈ/ is phonologically contrastive with the alveolar /d/. In English retroflexion does exist (compare the /d/ of "dill" [dɪll] with that of "drill" [ɖrɪl]) but is not phonological, thus although articulatorily and acoustically distinct, the two /d/s map to the same phoneme in English (i.e. they are allophonic variants). For this reason, it may represent a more difficult case than the /r/-/l/ distinction for Japanese. The trainees must learn to treat two phonetic categories as distinct that they have spent some 18+ years and countless iterations treating as identical.

Pruitt et al (2006) trained both NS of English and Japanese on perception of this contrast. The training was adaptive, in that initial training stimuli were truncated syllables (the vocalic portion was shortened), which has been shown to enhance NS of English perception of the contrast, and proceeded to full syllables. Also, subjects began with only one voice, and other voices were mixed in as they reached predetermined criterion levels. Training was fairly extensive – subjects were exposed to at least 8400 stimuli, though some of these were exposure only, and did not require a response by the subject. As predicted by the researchers, although both groups showed significant improvement in the test of generalization to a new voice, the Japanese group, for whom the contrast did not represent allophonic variants, showed significantly greater improvement. In post-training questionnaires, NS of English had apparently perceived the contrast as a difference in voicing, while NS of Japanese saw it as a place of articulation contrast. Further evidence of the difficulty of this contrast is provided by the aforementioned Sinnott et al (2007) study. They used the same methodology that successfully trained NS of Japanese to perceive the /r/-/l/ contrast as well as NS of English (at least for the training stimuli), but failed to find similar improvements for NS of English learning the retroflex contrast. Further research with other contrasts that are allophonic in L1, but represent different phonemes in L2 is called for and it would be particularly interesting to determine if adaptive training (with actual manipulation of acoustic parameters, rather than mere truncation of vowels) would help to guide learners' attention to the acoustic cues relevant to the discrimination.

Phonetic training studies have also been used to improve L2 perception of vowels (Wang & Munro, 2004; Iverson & Evans, 2009; Nishi & Kewley-Port 2007, 2008), tones (Wang et al, 1999; Wang, Jongman & Soreno, 2000) and quantity contrasts (Hirata, Whitehurst & Cullings, 2007; Tajima, Kato, Rothwell, Akahane-Yamada & Munhall, 2008; Menning et al, 2002; Tajima, Rothwell & Munhall, 2002; Hirata; 2004; Heeren & Schouten, 2008; Motohashi-Saigon & Hardison, 2009). The methodologies and findings of these studies are not profoundly different than those mentioned about for consonant contrasts, and they will not be discussed here.

4.3 Conclusion

This chapter reviewed the main results that have been obtained with phonetic training studies. As we have seen these studies have been used to improve the ability of many language learners to perceive speech contrasts in their L2 that they previously could not. The most successful studies have been those based on the methodology developed by Logan and his colleagues for training NS of Japanese to perceive the

English /ɹ/-/l/ contrast. This methodology involves the use of real-speech tokens from several different speakers, an identification task, and immediate feedback. As noted, we have less information about training studies that have failed to produce results.

Few of these training studies have addressed the pedagogical potential of phonetic training, and none have considered NS of English learning Arabic, despite the well-known difficulties that these learners face in accurately perceiving Arabic speech. In Chapter 5 we describe a PT study conducted with NS of English that attempts to train them to better perceive the /h/-/ħ/ contrast found in Arabic. Before that though, we first describe important aspects of Arabic phonetics and phonology to better understand the task confronting out learners.

Chapter 5: Empirical Investigation of Phonetic Training for Learners of Arabic

5.1 Introduction

In this chapter we will describe the results of a phonetic training experiment conducted with a group of Arabic learners. The main purpose of this experiment is to assess the viability of phonetic training as a pedagogical tool for adult learners of Arabic in a university setting. At a minimum, this requires that we demonstrate empirically measurable improvements in a perceptual task. Two important considerations motivated the experimental design. First, that the phonetic training implement the principles of high-variability phonetic training, which has proven effective in yielding perceptual improvements in learners of a number of different languages. Second that the demands placed on the participants be reasonable for language learners in a university setting that places many other demands upon their time.

5.1.1 Participants

The participants for this experiment were recruited from four different universities in the United States. Each of the universities chosen for participant recruitment had large Arabic programs (100+ students) with at least five contact hours per week in their first and second year classes. Participants were recruited through their instructors who were asked by the researcher to send flyers to their students describing the research. The criteria for participation was that the participants be currently enrolled in a second or third year Arabic class, that they be native speakers of English, that they not be heritage learners of Arabic and that they had never participated in a previous phonetic training study. Students were aware of the purpose of the experiment at recruitment.

Students wishing to participate contacted the researcher via email to request more information about the study. They were informed that the experiment would involve a pretest, possible assignment to phonetic training, which would take about 10 hours of their time over 4 weeks, and a posttest. Participants were paid $5 for the pretest, $75 for completing the training as required, and $5 for the posttest.

Forty-six individuals requested access to the pretest site and of those 41 (28 females) completed the pretest. Of those, 28 participants were in 2nd year classes and 13 in 3rd year. The average age of participants was 20.3 years (s.d =2.02, range 18-28).

Prior to taking the pretest, participants were asked questions about their language background to assure they were indeed qualified to participate. They were also asked a series of questions designed to assess their motivation from which a motivation score was compiled, and they were asked to rate their confidence in their ability to perceive the difference between Arabic /h/ and /ħ/. See Appendix A for documents connected with the perception test.

5.1.2 Stimuli

The 108 stimuli for the pretest were recorded from two native speakers of Arabic both of whom read a list of words containing minimal pairs of /h/ and /ħ/ as well as words containing a number of other phonemes. Stimuli were recorded with a Marantz PMD 670 solid-state recorder in a sound proofed booth in the UT Sound Lab (Calhoun 518). The words chosen were gleaned from textbooks for teaching the Arabic writing system (Brustad, Al-Batal, Al-Tonsi, 2004; Alosh, 2009) and from a leading Arabic-English dictionary (Wehr & Cowan, 1993). A sound file was created from each of these words and all of these files were equalized for volume. The minimal pairs featured the contrast in four different phonetic contexts – word initial (IN), intervocalic (IV), between a vowel and consonant (VC) and word final (FI). The words from the male voice were referred to as the training voice (TV) since stimuli from this voice were also part of the phonetic training to be described later, and the female voice was referred to as the generalization voice (GV) since this voice was heard by participants only at pre- and posttest and was used to determine if the gains from phonetic training would transfer to voices that were not part of training. TV stimuli were provided by a 25-year-old Palestinian male and GV stimuli by a 21-year-old Syrian female. The written words included were fully voweled, meaning that short vowels were marked for the participants. When texts are fully voweled in Arabic the orthography is shallow and words are pronounced exactly how they are spelled. Thus the minimal pairs were spelled exactly the same except for a single letter representing either /h/ or /ħ/. See Appendix A for a complete list of the test stimuli.

5.1.3 Pretest

Participants took the pretest online. They were provided with a username and password for a website hosting the pretest. Upon signing onto the website, participants were presented with an IRB-approved consent form which they consented to by clicking a link at the bottom of the page. Participants then took a pretest survey. Before the test began participants were asked to check boxes affirming that they were using the Firefox browser (other browsers had shown problems with the test site in pilot testing), that they were using a computer (rather than a tablet or smartphone), that they were wearing headphones, that they had turned off their mobile phone and that they would have 15 minutes of distraction free time to complete the test. Upon agreeing to these conditions, participants were given a short sample test with feedback. Sample items did not contain the /h/-/ħ/ contrast and the sample voice (the experimenter's) was not used in the pretest.

The test was a lexical identification task. Participants were presented with two words in Arabic script that were minimal pairs of one another on the screen and heard one of those words spoken. The participants' task was to click on the word corresponding to the sound file. For the sample test only, participants received feedback on their answer and if the answer was incorrect the item was repeated. After correctly answering the three sample test items, participants proceeded to the main test. The main test was similar to the sample test, but participants did not receive feedback on their responses. Once the participants clicked their answer for an item the next item was presented. Participants heard a block of 54 items from the training voice (TV) as described above, after which they had an opportunity to take a break if they desired. Then the next block of 54 items from the generalization voice (GV - also described above) was presented. Test items within a block were presented in random order and the order of the blocks was counterbalanced across participants. Each pair of words was presented twice with the correct answer alternating. So within each block /h/ was the correct answer 27 times and /ħ/ the correct answer 27 times. The phonetic context breakdown of stimuli was as follows:

Context	TV Stimuli	GV Stimuli	Total
IN	20	20	40
IV	10	10	20
VC	12	12	24
FI	12	12	24
Total	54	54	108

Table 5.1: Test stimuli by phonetic context

5.1.4 Pretest results

Five participants that scored higher than 75% correct on the IN GV test items were removed from the study as their performance indicated they were near or at ceiling and were unlikely to show further improvement. The table below provides more information about the participants.

	Male	Female	Total
2nd Year Students	8	17	25
3rd Year Students	2	9	11
Total	10	26	36

Table 5.2: Description of participants

Because participants were aware of the purpose of the test at recruitment (to improve /h/-/ ħ/ perception) and because participants that performed well on the pretest were eliminated from the analysis, these results should not be considered representative of all learners of Arabic, but rather those that have difficulty with the /h/-/ħ/ contrast.

Table 5.2 shows the mean percent correct scores for the GV and TV voices separately averaged across all 36 participants at pretest. A paired samples t-test reveals that there are no significant differences in the two scores (t_{35}=-.327, p=.745). Therefore, further analysis of the pretest data will include total scores, combining data from the two voices.

	Mean % Correct	Standard Deviation
GV	64.0	10.2
TV	64.5	13.2

Table 5.3: Mean per cent correct scores for GV and TV

In the pretest survey, participants were asked to rate their confidence in discriminating between /h/ and /ħ/ on a scale between 1 and 7 with 1 meaning they had no idea how to tell them apart and 7 that they had no difficulty telling them apart. The average rating for all 36 participants was 3.83 (sd=1.13, range 2-5). Given that participants were recruited based on the possibility of improving their ability to perceive this difference, it is not surprising that they did not express much confidence in that ability. A Pearson correlation test of pretest self-rating and pretest performance showed a positive correlation of r =. 26, (p = .031) indicating a statistically significant, but small, relationship between confidence rating and score.

It is also of interest to know if participants performed significantly better on any one phonetic context compared to the others. The table below shows descriptive statistics for all 36 participants on each individual phonetic context. Numbers shown are per cent correct scores.

Context	Minimum	Maximum	Mean	Std. Deviation
IN	35.0	80.0	58.0	11.2
IV	5.0	95.0	59.9	19.7
VC	29.2	91.7	65.4	14.3
FI	29.1	100	77.3	14.8

Table 5.4: Descriptive statistics of pretest results, by context (% correct)

Boxplots of these data are shown below. The circles, labeled by subject number, indicate outliers. These points are those that lie below Q_1-IQR*1.5 or above Q_3+IQR*1.5 where IQR is the interquartile range, Q_1 is the measure of the first quartile and Q_3 is the measure of the third quartile.

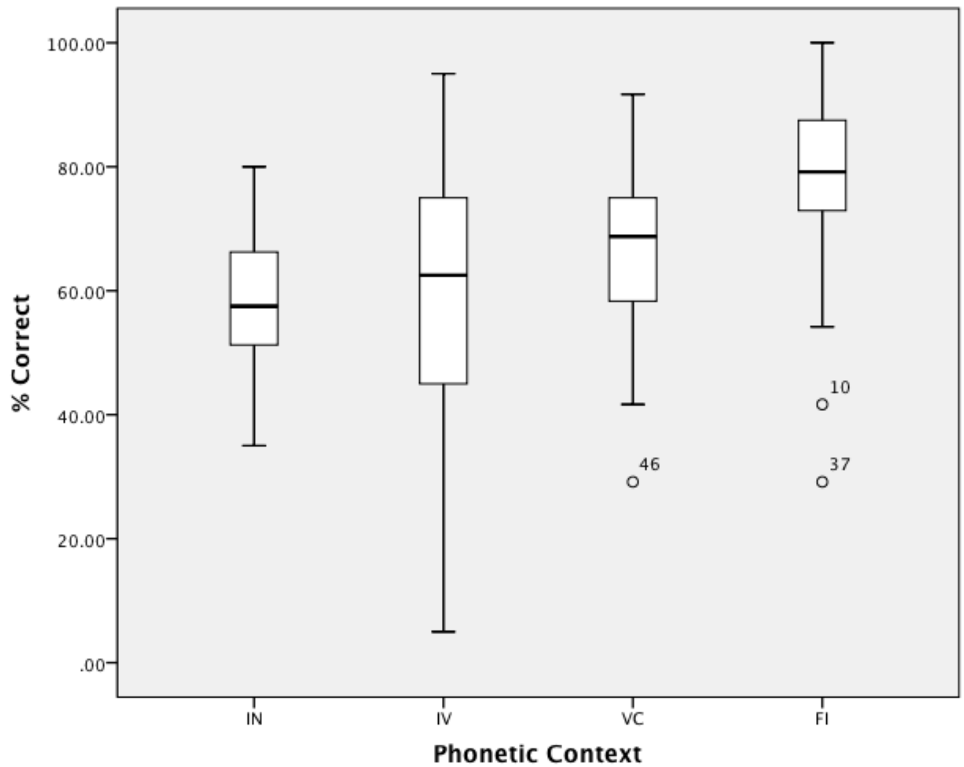


Figure 5.1: Pretest results by phonetic context

The table below shows results of a Bonferroni corrected ANOVA comparing pretest scores for each context with one another. Significant differences are starred.

Context		Mean Difference	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
IN	IV	-.019	.036	1.00	-.115	.078
	VC	-.074	.036	.251	-.171	.022
	FI	-.193	.036	.000*	-.290	-.097
IV	VC	-.055	.036	.764	-.152	.0412
	FI	-.175	.036	.000*	-.271	-.078
VC	FI	-.119	.036	.007*	-.216	-.022

Table 5.5: Pairwise comparisons of phonetic context scores

These results show that the FI context was significantly easier for the participants than the other three contrasts, but that performance on the other three contexts was not significantly different. It is also interesting to note the substantial variability we see in performance on the IV context compared to other contexts. The presence of very low scores suggests that some participants were mislabeling the phonemes in that context.

The above analysis and discussion utilizes percent correct scores as the dependent variable because it is the easiest measure of subject performance to understand. However, in a situation where only two possible response outcomes are possible (i.e. /h/ or /ħ/) we must consider the possibility that subjects are biased in favor of one response or the other and that this bias may influence the results that we obtain. Therefore, we will now discuss a more complicated statistical measure of sensitivity that may give a more accurate picture of our participants' sensitivity to the contrast in question. Two statistical measures of participant performance will be calculated. They are β (beta) and d' (d prime). β is a measure of bias and will be used to determine if participants were more likely to respond /h/ or /ħ/ and d' is a measure of sensitivity and attempts to give a bias free measure of participants ability to make the discrimination. Since the pharyngeal /ħ/ is, for native speakers of English, a novel articulation we want to know if our participants were able to detect this novel pharyngeal feature when it was present, and correspondingly if they report detecting it when it is not in fact present. To do this we need to calculate the percentage of times that participants correctly detect the existence of /ħ/, called the hit rate (H), and the percentage of times they report hearing /ħ/ when it is not in fact present, called the false alarm rate (F). To calculate d' we take z-scores of each of these values. Z-scores are a measure of the number of standard deviations away from the mean that a given date point is. The mean is set to 0 and the standard deviation set to 1 for this calculation. For a given participant d' = Z(H)-Z(F). This is essentially the ratio between two different distributions - signal and signal + noise (Perlman, 1990). Although the value of d' is theoretically unlimited (100% correct, with no false alarms) in practice they range up to 2.0 (Keating, 2005). To calculate β we use $\beta = \exp(-1*((Z(H)^2 - Z(F)^2)*.5$. An unbiased observer will have a β score of around 1.0. As bias towards detection increases (responding /ħ/ in our case) β will approach 0 and as bias against detection increases (responding /h/ in our case) β will increase. We now present analysis of our pretest data using β. d' will be more relevant later when we compare our training group with a control group.

	Minimum	Maximum	Mean	Std. Deviation
β	.58	1.30	.921	.166

Table 5.6: Pretest β scores (N=36)

These data show that our participants were biased slightly in favor of detecting the pharyngeal /ħ/ articulation. A single sample t-test shows that the average β (.921) is significantly different from 1.0 ($t_{35} = -2.85$, $p = .007$). It is not entirely clear why this should be since it would seem that the pharyngeal /ħ/ is novel acoustically and articulatorily for NS of English. Examining β scores for each different context though, appears to provide an answer.

Context	Minimum	Maximum	Mean	Std. Deviation
IN	.72	3.59	1.13	.514
IV	.07	6.58	1.05	1.41
VC	.07	4.08	1.10	.611
FI	.07	2.80	.44	.547

Table 5.7: Pretest β scores by phonetic context (N=36)

Notice first that there is substantial variation in the bias found for each phonetic context. For each there were participants biased towards /h/ and participants biased towards /ħ/. However, on average subjects showed little bias in either direction for the

IN, IV and VC contexts, but a strong bias towards /ħ/ in the FI context. It is difficult to be certain as to why this may be, but it seems quite possible that this is due to the fact that English does not have word final /h/, so for example the word "yeah" is pronounced /jeə/ not */jeəh/. Thus it may be that participants are categorizing any word final aspiration as the novel /ħ/.

5.2 Phonetic training

Following the pretest, two groups of 18 participants were formed. Participants were pseudorandomly assigned to either the control group or the training group. To the extent possible the two groups were created so that their average pretest performance was more or less identical and so that there was balance between the groups in terms of the source institution and current training level. So, for example, if there were 11 participants from institution A, 5 were assigned to one group and 6 to the other.

Participants assigned to the control group were informed that they would be asked to return in about 4 weeks time to take a posttest identical to the one that they just took. Participants assigned to the training group were given access to a phonetic training website and were asked to complete 100 training modules over about 4 weeks.

The choice of a training website, rather than a more easily controlled laboratory setting was based on the desire to make the training easy to use and to provide flexibility for the trainees who were able to access the site at any time and from any place with internet access. Furthermore, since this experiment was designed to evaluate the efficacy of phonetic training as a pedagogical tool, it was deemed best to test that tool in the same way that it would eventually be used by learners of Arabic. Participants were encouraged to keep a regular training schedule of 5-8 modules a day, but were able to use the site according to their own schedule.

The website was simple to use. Participants would log onto to the website and be taken to a welcome screen (see Appendix B for screen shots of the training website). This screen showed them how many modules they had completed, the number remaining to completion, the last time they used the website and the highest score they had obtained on a module. When the participant was ready to complete a module they would click on a button marked “Begin Training” and the test would begin. Each test item consisted of a test screen and a feedback screen. On the test screen, participants would see two words written in Arabic script on their screen. In between them was a button that participants clicked to hear one of those words spoken. They could click the button as many times as they wished and when they were ready to respond they clicked on the word corresponding to the word they were hearing. This would bring them to the feedback screen. The feedback screen first indicated to them whether or not they had answered correctly and also provided several other listening opportunities. The participant could choose to hear the same word again, or to hear the word’s counterpart from the same voice. The participant could also hear either of those two words from four additional voices. Each training module has 12 different pairs and each was presented twice per module with the correct answer alternating. There were 5 different modules, each provided by a different voice. All stimuli on the training website featured the target contrast in word initial position. As indicated, participants could hear any of the five voices producing a given pair at the feedback screen, though only the voice for that module was available at the test screen. When the participants completed a given module they were returned to the home screen where they could choose to begin another module or log out. Participants also had the option to view a progress chart from the home screen that showed graphs of their progress over time on the modules, broken up by voice. One of the voices used in the training was the same as the TV voice indicated in the pretest while the other 4 voices were heard only during the training sessions. Recall that the GV voice was heard by subjects only at pretest and posttest and never during training. The ages, sexes and native dialects of the 5 voices are indicated in the table below:

	Age	Sex	Dialect
Voice 1 (TV)	25	Male	Jordanian
Voice 2	38	Female	Lebanese
Voice 3	34	Male	Palestinian
Voice 4	34	Male	Egyptian
Voice 5	22	Female	Palestinian

Table 5.8: Description of training stimuli voices

After participants completed the required 100 training modules (each of the 5 different modules presented 20 times), they were contacted for a posttest.

5.3 RESULTS

The posttest was identical to the pretest.

Only 10 participants from the training group completed all 100 training modules and returned for the posttest and only 14 participants from the control group returned for the posttest.

As indicated in Chapter 4, the most important test of the effectiveness of training from a pedagogical viewpoint is the test of generalization. Since the participants only heard this voice at pre- and posttest, and since they never received feedback on their identification of /h/ and /ħ/ with this voice, it is the best measure of the extent to which participants' new perceptual abilities are language general, rather than stimulus specific. The simplest test of training gains is to compare the pretest and posttest scores of participants in the training group. Table 5.9 below shows mean per cent corrects scores and standard deviations at pretest and posttest for the 10 subjects that completed the training modules.

	Minimum	Maximum	Mean	Standard Deviation
Pretest	38.9	81.5	65.4	11.8
Posttest	63.0	88.9	80.6	8.3

Table 5.9: Training Group (N=10) – Descriptive statistics for pre- and posttest - % correct

A paired-samples t-test comparing pre- and posttest scores for participants in the training group shows that the average improvement of 15.2% was statistically significant at the α = .05 level (t_9=-4.65, p=.001). However, it is possible that this improvement is the result of task familiarity (they took the same test twice) or the fact that they were concurrently enrolled in intensive Arabic language classes and the improvement is the result of that and not of the training website per se. Therefore, it is necessary to compare the difference in pre- and posttest performance in the training group with that of a control group that did not undergo training.

If training participants showed greater improvement than the control participants from pretest to posttest as measured by the test of generalization it suggests that they have gained knowledge about the difference between /h/ and /ħ/ that successfully transfers to new voices.

Below are descriptive statistics of percent correct scores on the test of generalization for the control group at pre- and posttest:

	Minimum	Maximum	Mean	Standard Deviation
Pretest	42.6	79.6	64.6	9.98
Posttest	48.1	83.3	68.4	11.7

Table 5.10: Control Group (N=14) – Descriptive statistics for pre- and posttest - % correct

The average improvement of 3.8% was not significant at the $\alpha = .05$ level for the control group ($t_{13} = -1.72$, $p = .109$).

Note that although not all participants returned, the average pretest scores for the training (65.4) and control (64.6) groups were similar and the difference between them was not statistically significant ($t_{22} = .184$, $p = .856$). We are fortunate that despite the loss of so many participants, the pretest performance of the two groups remained nearly equal.

A 2 x 2 repeated-measures ANOVA was performed to test for an interaction between TIME (pretest and posttest) and GROUP (training and control) on the 54 item GV test to determine if the change in scores between pre- and posttest was significantly different in the training group as compared to the control group. Boxplots, ANOVA results, and a brief description of the data follow below.

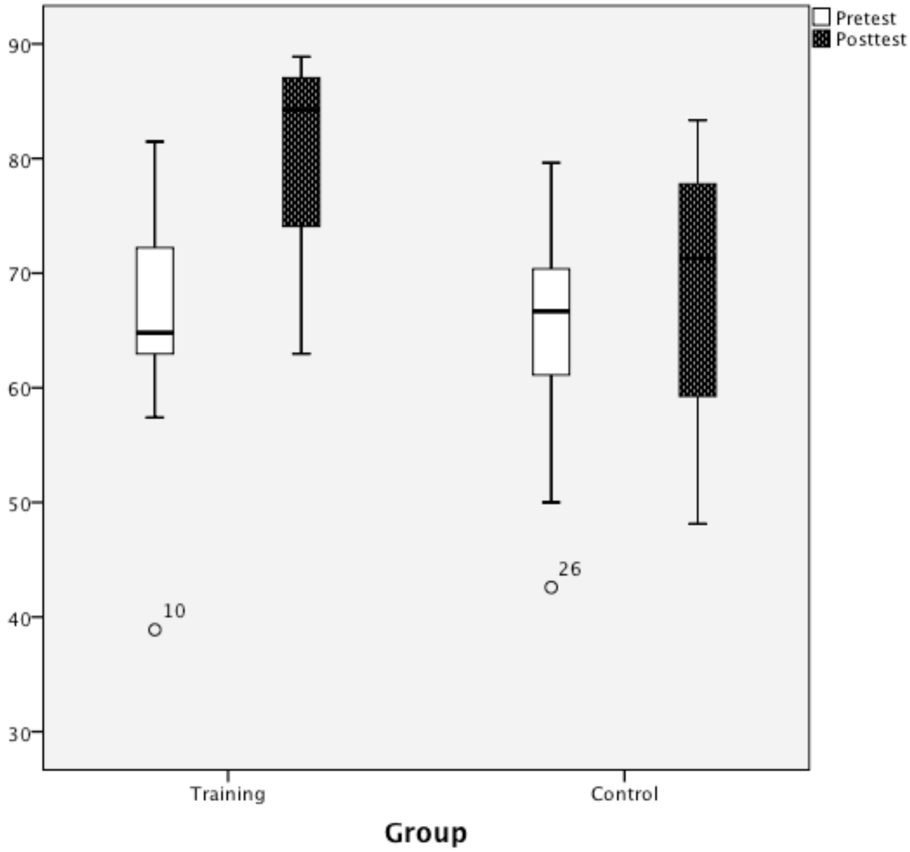


Figure 5.2: Generalization test results - % correct

	Type III Sum of Squares	Df	Mean Square	F	Sig.	Partial eta squared
Time	1055.26	1	1055.26	24.956	.000	.531
Time*Group	375.68	1	375.68	8.885	.007	.288
Error(Time)	930.26	22	42.285			

Table 5.11: Generalization test - % Correct – RM ANOVA results

Levene's test reveals that the differences in the variance of the pre- ($F_{1,22}$ = .087) and posttest ($F_{1,22}$ = 1.86, p = .186) scores were not significant and so uncorrected results are reported here. The test shows that the greater improvement shown by the training group is statistically significant ($F_{1,22}$=8.89, p = .007) with a medium effect size

(η_2 = .288). This is the most important statistical result of the experiment as it demonstrates that the training group showed significantly greater improvement than the control group in perceiving stimuli provided by a voice with which the two groups had equal experience. As noted, it is necessary, if not sufficient, to be able to show that phonetic training can yield language general improvement in the ability to identify /h/ and /ħ/ and we have done that here.

As discussed in section 5.1.4, d' (d prime) is a bias free measure of sensitivity and as such it is worth considering if improvements in d' for the training group were, like the improvements in per cent correct scores, significantly better than those of the control group. Below we show descriptive statistics for both groups at pre- and posttest measured in d' and the results of a RM ANOVA testing for an interaction between TIME and GROUP as above.

Test	Group	Mean d'	Std. Deviation	N
Pretest	Training	.795	.653	10
	Control	.848	.570	14
Posttest	Training	1.842	.606	10
	Control	1.030	.666	14

Table 5.12: Descriptive Statistics – d' scores

	Type III Sum of Squares	Df	Mean Square	F	Sig.	Partial eta squared
Time	4.410	1	4.410	25.651	.000	.538
Time*Group	2.179	1	2.179	12.673	.002	.365
Error(Time)	3.782	22	.172			

Table 5.13: Generalization test – d’ scores – RM ANOVA results

Comparing pre- and posttest d’ scores lends further evidence that participants have become more sensitive to the /h/-/ħ/ discrimination. Subjects in the training group had an average d’ score of 1.842 after training indicating greater sensitivity and reduced bias in detecting the pharyngeal articulation for /ħ/. The effect size using d’ scores (η_2 = .365) is even greater than that found using per cent correct scores (η_2 = .288).

Since the group sizes are rather small, it is worth looking at results for each of the individuals in the study. The next two figures show pre- and posttest scores for both the control and training groups.

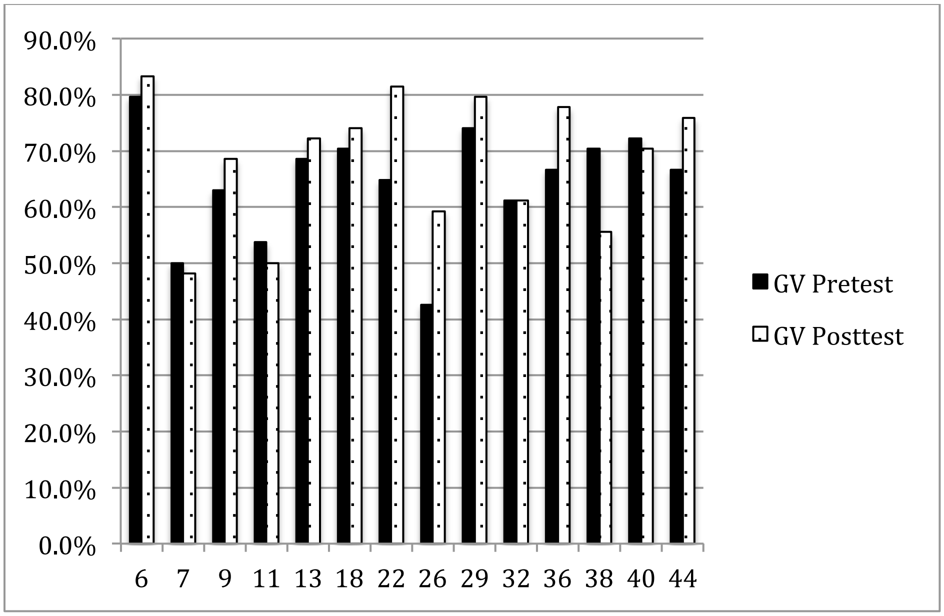


Figure 5.3: Generalization Test: Individual Pre- and Posttest Scores – Control Participants

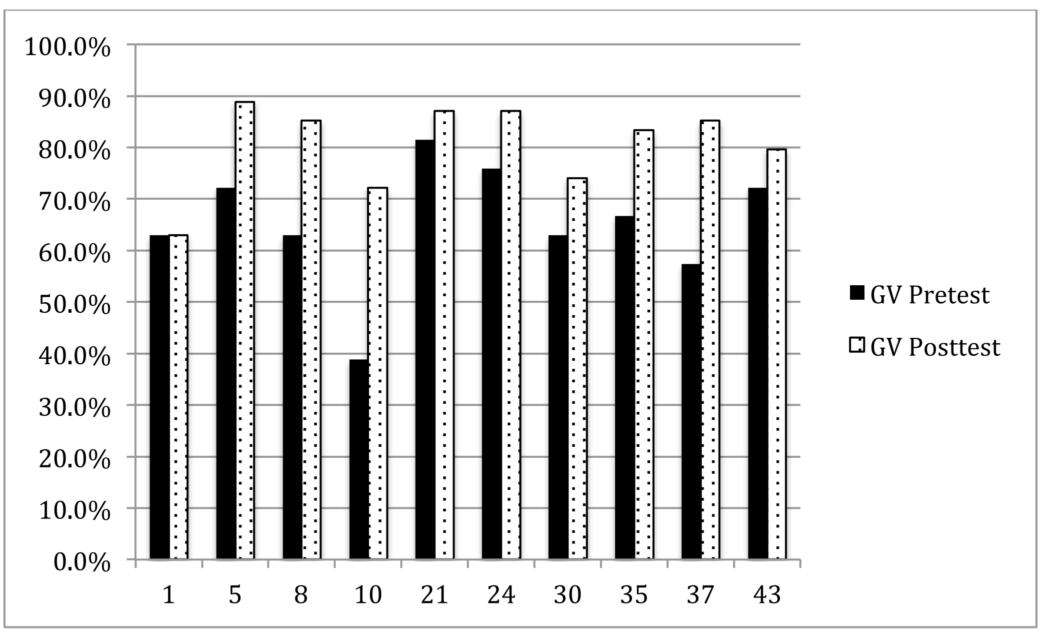


Figure 5.4: Generalization Test: Individual Pre- and Posttest Scores – Training Participants

Note that while as a group the trainees showed greater improvement than the control group it is not the case that all trainees showed greater improvement than all members of the control group. In fact one trainee showed no improvement from pre- to posttest. This is not an unexpected result. As we saw in Chapter 4, the gains of phonetic training tend to be spread unevenly across participants.

It is important, in light of the completion rate of training (only 56% of participants assigned to the training group completed training), to consider the reason that so many of the trainees failed to complete the training. If they dropped out because they felt that the training was not working, substantial bias would be introduced into the final results. Only 2 of 8 dropouts gave reasons for discontinuing the study. One indicated that medical problems prevented her continued participation, and another that her workload at school forced her to dropout. Of the remainder, most completed fewer than 5 modules and none more than 23. We conclude then that the dropout rate, while high, is most likely due to the demands that training placed on participants time, rather than the perception that training was ineffective.

A further potential confound is that it was not possible to control precisely for time between pre- and posttest since participants had some degree of freedom in choosing when to take the tests and how long it took to go through the training. In fact only a few participants finished the training in the requested 4 weeks and one took almost 8. It may be that improvement in the test scores in the training group can be explained by a lengthier (or shorter) interval between pre- and posttest. The tables below show the interval between pre- and posttest for all participants and the improvement score (posttest minus pretest) for the generalization test for the training and control groups.

Participant Number	Pre/Posttest Interval (days)	Improvement Score
6	35	3.70
7	35	-1.85
9	45	5.56
11	35	-3.70
13	27	3.70
18	39	3.70
22	40	16.67
26	35	16.67
29	47	5.56
32	31	.00
36	43	11.11
38	35	-14.81
40	44	-1.85
44	32	9.26
Average	37.36	3.84

Table 5.14: Control Group - Pre/Posttest interval and improvement scores

Participant Number	Pre/Posttest Interval (days)	Improvement Score
1	35	.00
5	48	16.67
8	33	22.22
10	55	33.33
21	34	5.56
24	29	11.11
30	33	1.11
35	31	16.67
37	32	27.78
43	27	7.41
Average	35.7	15.19

Table 5.15: Training Group - Pre/Posttest interval and improvement scores

There was no significant difference between the average interval between pre- and posttest of the two groups (t_{22} = -.556, p = .584) and a Pearson's test revealed no significant correlation between the interval between tests and the improvement score (p = .246). Therefore it does not seem that the improvement shown by the training group is the result of a longer or shorter interval between pre- and posttest.

While the training itself used only stimuli featuring the contrast in the word initial position, the pre- and posttests included stimuli from other phonetic contexts as well, and thus another important question to be addressed is whether or not training improvements generalized to other contexts. To address this a 2 x 2 x 4 RM ANOVA was conducted with the variables GROUP, TIME and CONTEXT respectively. Using a Bonferroni correction, this test fails to find a significant Time*Context*Group interaction ($F_{2.24, 87.03}$ = .639, p = .55); however, due to the multiple levels of interaction and a relatively small data set, the observed power of the test is only .157 meaning that it is unlikely to reveal any relationships that do exist. Finding meaningful interactions at the level of phonetic context will most likely require larger participant groups and lengthier assessment tests.

Curiously, the results for the training voice, which the training participants heard and received feedback on, do not show a significantly greater improvement for the training compared to the control group, however the result comes very close to reaching statistical significance, as seen below.

Test	Group	Mean d'	Std. Deviation	N
Pretest	Training	66.48	15.257	10
	Control	62.70	12.195	14
Posttest	Training	84.26	6.250	10
	Control	71.16	14.631	14

Table 5.16: Training test – Descriptive statistics – % correct

	Type III Sum of Squares	Df	Mean Square	F	Sig.	Partial eta squared
Time	2008.76	1	2008.76	33.48	.000	.603
Time*Group	252.9	1	252.9	4.22	.052	.161
Error (Time)	1320.0	22	60.00			

Table 5.17: Training test – RM ANOVA results - % correct

Thus while the results on the generalization test show significant improvement for the training group, the training test results fall just shy of significance at the $\alpha = .05$ level ($F_{1,22}$=4.22, p = .052). It is peculiar that this should be the case since the training group had exposure to at least 480 tokens of /h/ or /ħ/ from the training test voice, with feedback. This may be due in part to the fact that the groups were not equal at pretest for the training voice. The control group averaged 62.7% while the training group averaged 66.5% meaning that the control group had more room for improvement. Furthermore, posttest results for the training group show less variation (s.d. = 6.3) and a Levene's test for equal variances suggests that the scores at posttest do not have equal variance ($F_{1,22}$ = 15.7, p = .001). It may be then that there is a ceiling to how well learners can be expected to do on this particular test and that some of the trainees reached that ceiling for the training voice test.

In summary, the results of the training experiment are encouraging for our purposes. They show that, in comparison with no training, HVPT can improve perception of the /h/- /ħ/ contrast in learners of Arabic even when the constraints of the training are relaxed to make it more useful for our learners. Critically, there is evidence that the perceptual gains of training are language general, rather than stimulus specific, and although the experiment sacrificed some rigor, introducing the confounds of participant loss and varying intervals between pre- and posttest, we can be reasonably confident in assigning some of the improvements seen to phonetic training.

5.4 Training data

In this section, we review the data provided by the training website to look for patterns in learning rate and any difference that may have existed in performance on the five different voices.

Each of the training modules had 24 test items. The figures below show the percent correct scores for each of the participants for training modules. Rather than showing the scores for individual modules, the charts show the scores for groups of ten modules. So 1st is the first 10 modules, 2nd the next ten and so on. This eliminates the volatility introduced by differential performance on the different training voices, which will be treated below.

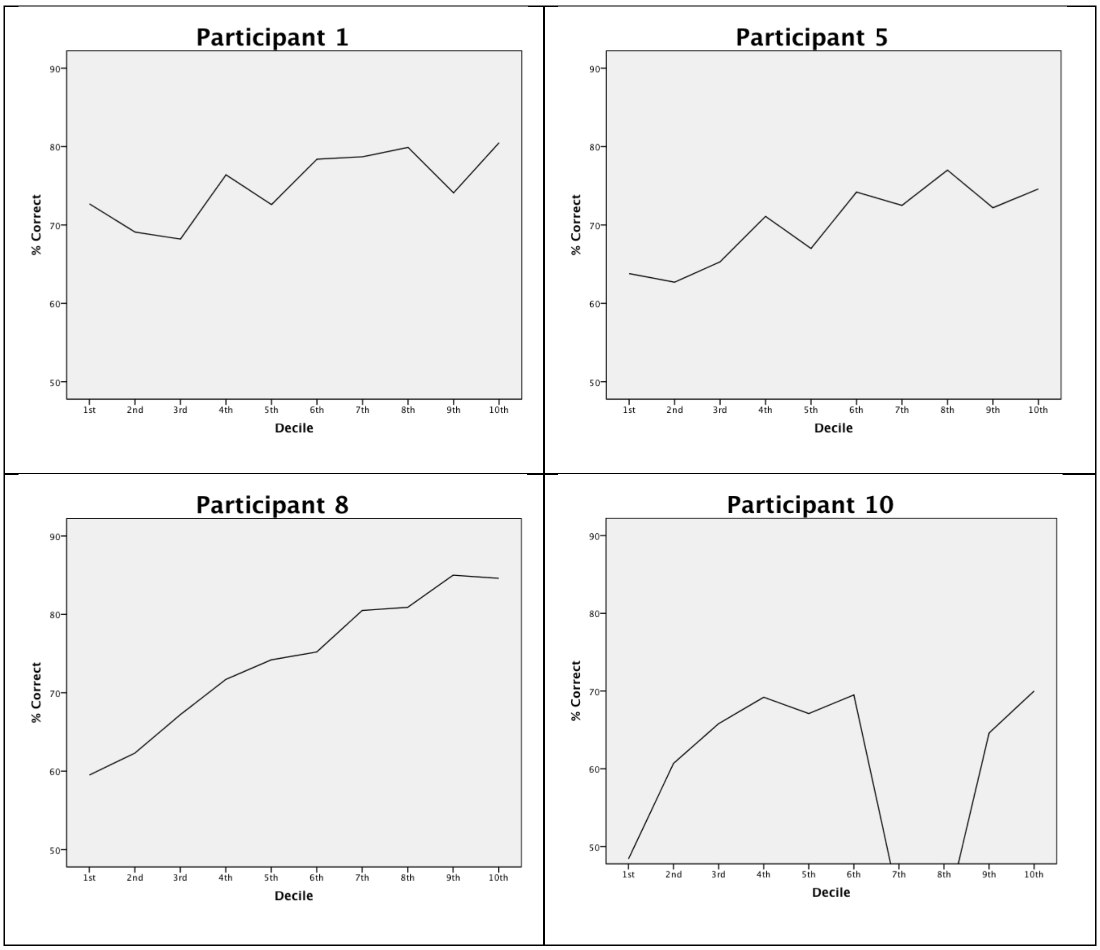


Figure 5.5: Individual training results over time

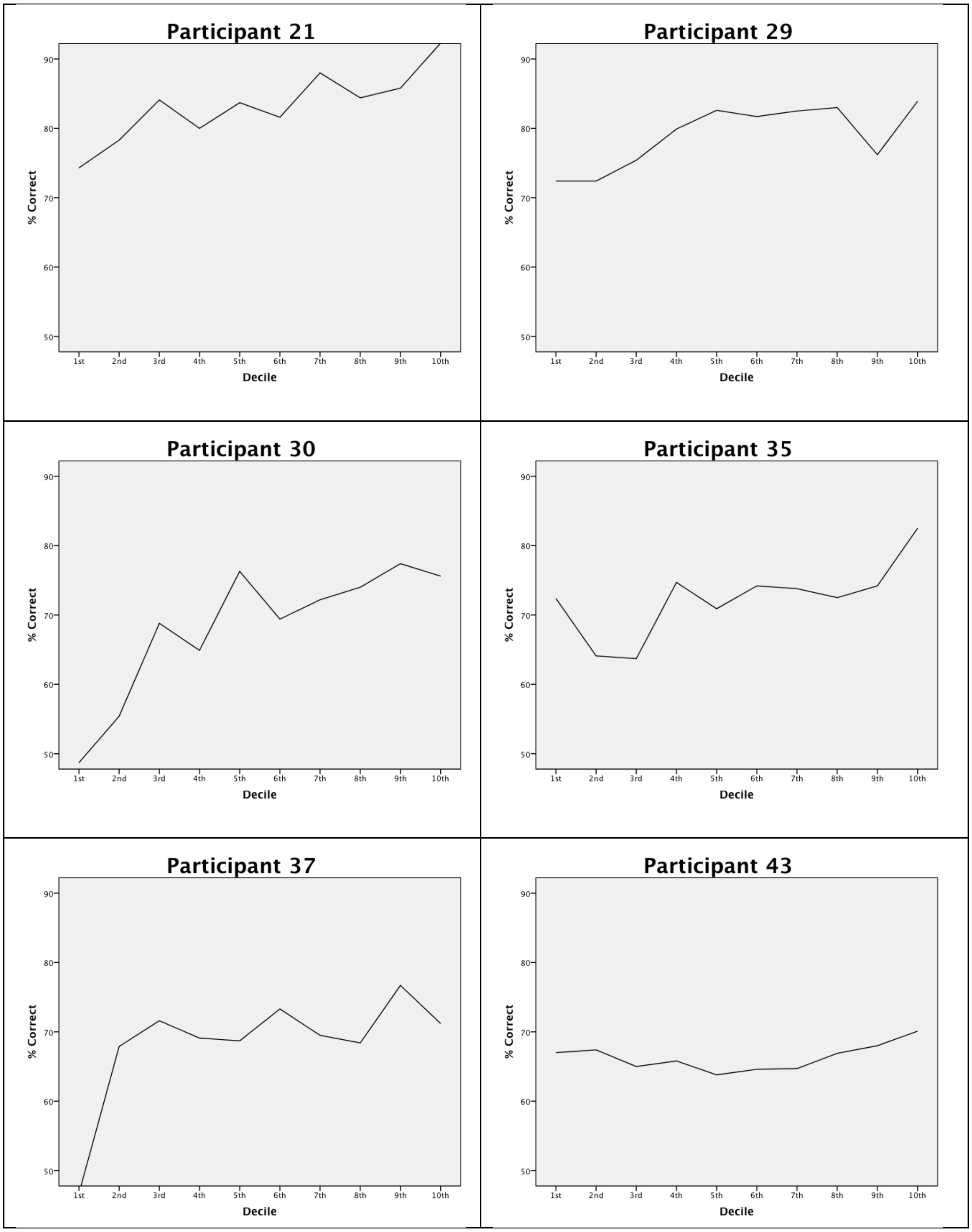


Figure 5.5, cont.: Individual training results over time

No consistent patterns emerge in examining the individual progress charts. Participant 8 shows the most steady progress over time consistent with his improvement on the generalization test (63% to 85% correct). Similarly, Participant 43 shows little improvement over the course of training, and little difference between pre- and posttest generalization test scores (72% to 79%), though this may also be due to ceiling effects. Participant 10 shows a very peculiar result. It seems that at some point this participant may have reversed his category assignment as his scores drop as low as 20% for some modules. He appears eventually to have recognized the error as training scores in the final modules improved and the participant showed significant improvement at posttest (38% to 72%). Indeed it may be that this participant, and perhaps others, had some ability to perceive the contrast, but did not know which was which. Participant 37 seems to show rapid progress early, and then a leveling out, also suggestive of 'label learning'. To separate out perceptual learning from label learning would require a third participant group that underwent low variability training (using only one voice for example) to compare with the group undergoing high-variability training.

The figure below shows the average score of all participants for each group of ten modules

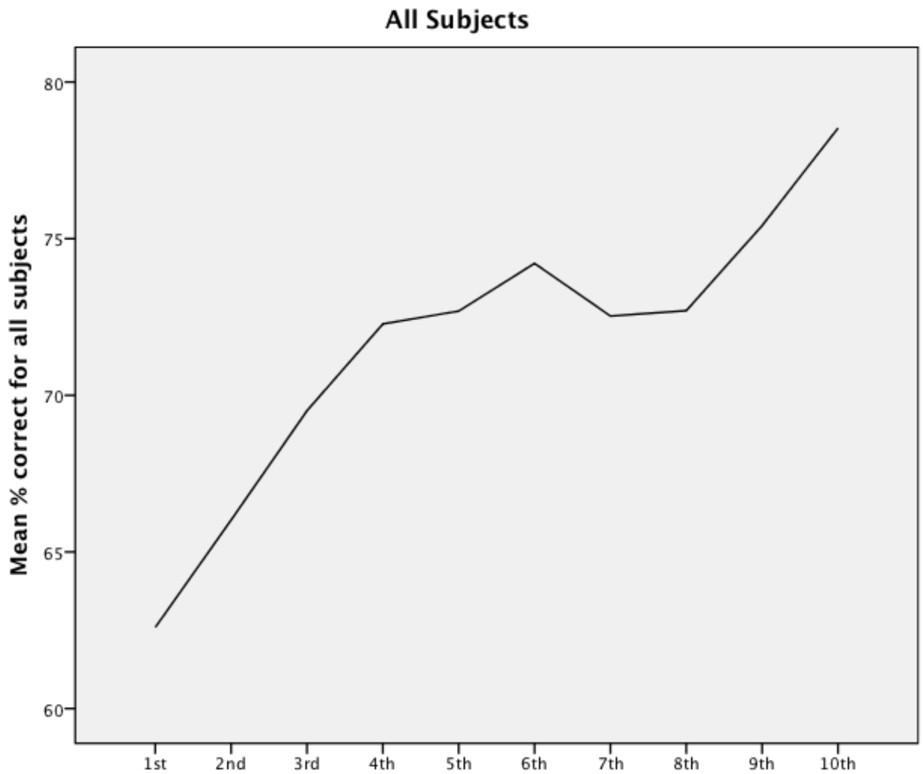


Figure 5.6: Training results over time, all participants

As can be seen from the charts above, this average line conceals substantial interparticipant variability, but also suggests steady improvements for the group. An ANOVA shows that there were significant differences in trainee performance across time ($F2 = 8.78$, $p=.001$). The table below shows Bonferroni corrected post hoc tests comparing average % correct scores for the 1st, 5th and 10th group of modules. Significant differences ($\alpha = .05$) are starred and show that subject showed significant improvement from the 1st to 5th group of modules, but the improvement shown between the 5th and 10th groups was not significant. This suggests that at least for the training stimuli improvements were greatest in the early part of training. It is not certain if the same would be true of the generalization stimuli which were only tested twice.

Time		Mean Difference	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1st	5th	-10.1	3.85	.043*	-19.93	.078
	10th	-15.9	3.85	.001*	-25.78	-6.11
5th	10th	-5.84	3.85	.423	-15.67	3.99

Table 5.12: ANOVA of training progress, all participants

The figure below shows boxplots of average participant performance on the first and last four modules for each of the voices. An omnibus ANOVA ($F_{4,45}$=1.49, p=.223) finds no difference in overall participant performance on the different voices during the first four modules.

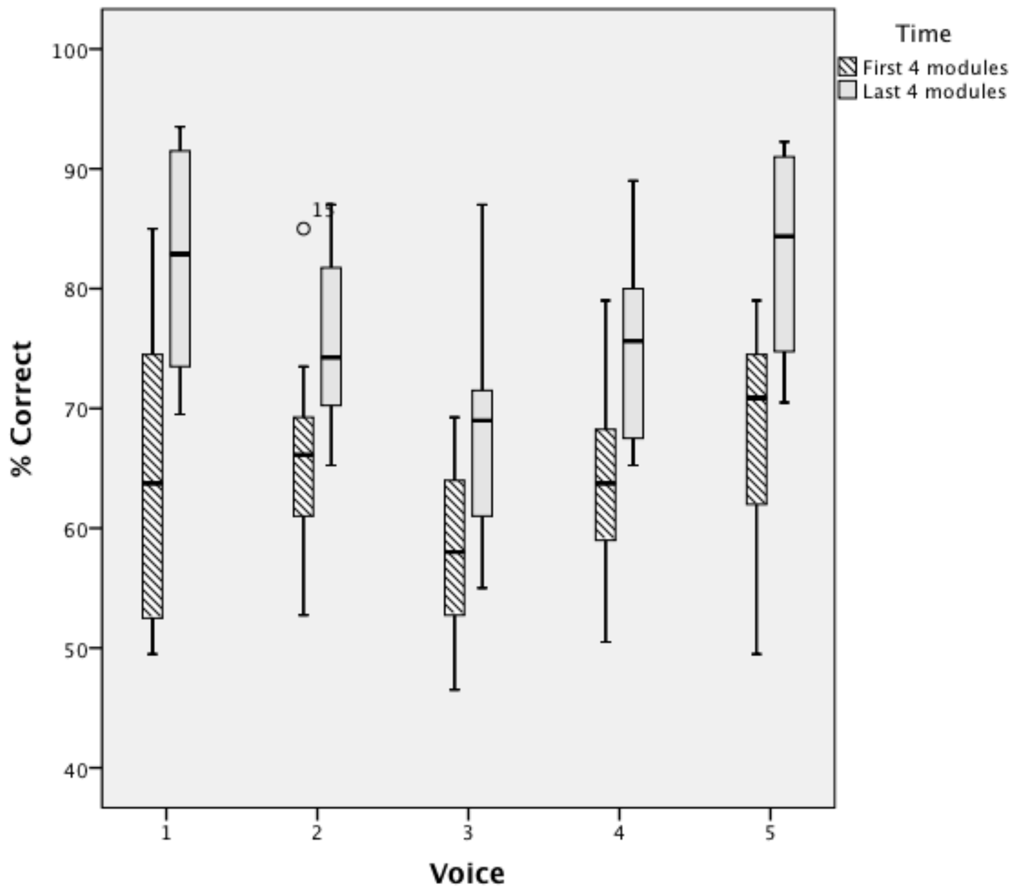


Figure 5.7: Training stimuli improvement, by voice

A Bonferroni-adjusted RM ANOVA with Voice and Time as factors reveals that as a whole participants showed statistically significant improvement on 4 of the 5 voices. The average improvement on Voice 3 proved not to be significant, however, it is uncertain if this is because Voice 3 was particularly difficult to learn, or if the test is simply not powerful enough to reveal statistically significant differences. The results of this ANOVA are shown below and significant results at the $\alpha = .05$ level are starred.

Voice	$F_{1,18}$ test	p=
1	12.93	.002*
2	6.19	.023*
3	7.22	.107
4	9.27	.007*
5	15.22	.001*

Table 5.13: ANOVA of training stimuli improvement, by voice

Thus, the participants found each of the voices more or less equally easy to learn with the possible exception of Voice 3.

5.5 PARTICIPANT FEEDBACK

In addition to the data described above, participants in the experiment were asked to rate the effectiveness of the training and to comment in general about their training experiences. Also, several of the training group participants consented to telephone interviews to discuss their experiences. It is intended from the surveys and interviews to find out if the participants felt that training was a valuable experience for them as language learners and to find out how the training website might be improved.

6.5.1 Survey results

The results of the survey will be summarized here. A complete copy of the survey can be found in Appendix A

Both the trainees and control group participants were asked to rate the extent to which their Arabic had improved overall, and more specifically whether their listening comprehension and pronunciation had improved. Participants were asked to rate their level of agreement with the statement "Overall I think that my Arabic has improved since taking the pretest" and their answers were converted to numerical scores from 0-5. On average participants that participated in the training scored a 4.25 on this question while the control group averaged 3.5, a significant difference (t_{22}=3.464, p=.002). The fact that the participants participated in some treatment may alone be responsible for this result as we may expect them to be inclined to perceive greater improvement from the fact of participation alone. Nevertheless the result suggests that trainees did feel that the training was of benefit to them as Arabic learners. This is important as potential trainees would be unlikely to persist through lengthy and no doubt tedious training if they do not feel that it is worth their time. For the specific question about improvement in listening comprehension the trainings average rating was 4.25 while the control group average was 3.19, again a significant difference (t_{22}=4.72, p=.001). When asked about pronunciation improvement the average score of the training group (3.50) was not significantly different from the average score of the control group (3.19) (t_{22}=1.048, p=.306). Therefore we do not find support for the idea that the gains of perceptual training may carry over to pronunciation.

Since we are, in the end, concerned with the extent to which phonetic training may improve Arabic language proficiency it is important to consider whether any gains from training transfer to real world language activities. Since it is difficult, if not impossible to empirically assess improvements in listening comprehension, it was hoped that trainees themselves could provide some insight as to what extent the training helped them in such tasks. Training participants were asked whether or not they agreed with statements about what circumstances they had improved their ability perceive the difference between the /h/ and /ħ/ in isolated words, in listening comprehension activities, and in conversations. Since there were only 10 trainees that finished the training, the data collected here can be suggestive only. Nevertheless, we present here participant ratings of the extent to which they agree with the statement "As a result of training, I can now tell a ح from ه in isolated words (also, while doing listening comprehension activities and during conversations in Arabic).

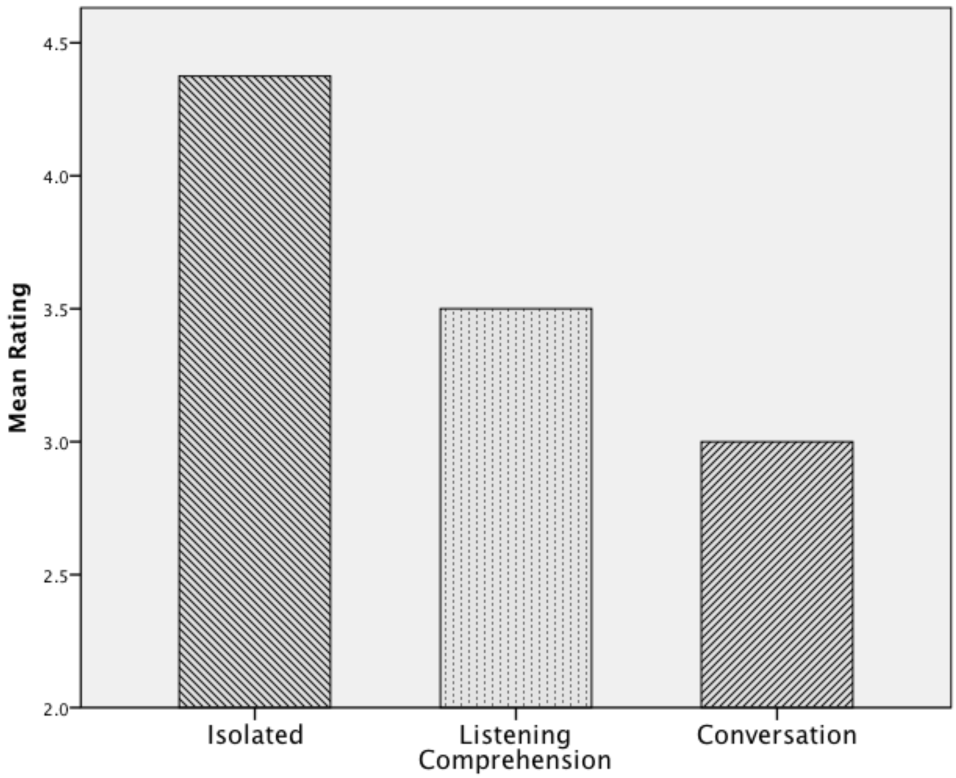


Figure 5.8: Trainees' mean ratings of their perceptual improvements in different tasks

An ANOVA reveals that there were significant differences in how participants rated their improvement depending on the required task (F_2=9.72, p=.002). A Bonferroni corrected pairwise comparison reveals that participants rated their training improvements in an isolated word task (mean rating = 4.38) significantly higher (p=.004) than improvements in the more challenging task of participation in a conversation (3.00). This result suggests that, at least initially, training gains in perception are more limited to simple tasks. It is important to determine whether or not these perceptual gains do eventually show up elsewhere, or if more explicit interventions are required to help learners incorporate these new perceptual abilities in tasks more closely approximating actual language use.

Finally, trainees were asked to indicate whether or not they felt that the training was worth the time they spent on it, and whether or not they would recommend it to a friend studying Arabic. Asked whether they agreed with the statement "the training was worth the time I spent on it" nearly all "Agreed" with one neutral and one "Strongly Agreeing" and the same was true in response to the statement "I would recommend the training to a friend studying Arabic." While it is gratifying that none disagreed with these questions about the value of the training, a more enthusiastic response would have been preferred.

Overall, the results remain encouraging for our research. In addition to statistically significant empirical results demonstrating perceptual improvements in the trainees that were not found in the control group, we have positive feedback from the trainees themselves regarding the value of the training to them. None of the participants in this admittedly small group felt that their time was wasted by the training.

5.5.2 Interview results

Only 5 of the trainees (Participants 5, 8, 10, 21, and 37) consented to participate in interviews about their experiences with the training website. In this section the feedback they gave will be summarized. See Appendix C for a list of the questions used in these interviews and for transcripts of four of the interviews (due to technical problems one of the interviews was not recorded). Participants were asked to comment on the usability and functionality of the website, on the extent to which they used the feedback and progress charts, if they developed any specific strategies for hearing the sounds, and if they felt the training was worth the time they spent on it.

None of the interview participants reported any significant technical issues with the website, though a couple indicated some delay in playing the test sound file occasionally. Most of the participants indicated that the site was easy to use and intuitive.

For each of the interview questions I summarize relevant comments from each participant.

1) Did you use the progress charts that the website provided?

Participant 5 – “I didn’t really find them that helpful because the trends didn’t really make sense to me.”

Participant 8 – “I used them a lot, every 10 lessons or so I would go check how I was doing.”

Participant 10 – Didn’t use them.

Participant 21 - “Progress wasn’t something that I was really concerned with.”

Participant 37 – “I really liked the progress graph . . . and I looked at it several times”

2) Did you use the available feedback options (i.e. to hear both versions of the word from multiple voices)?

Participant 5 – “For the most part I listened . . . to the current voice that I was using for the module.”

Participant 8 – “Towards the beginning I [used] it a lot, because I had no idea how to identify one from the other . . . towards the end I didn’t need to do it as much.”

Participant 10 – "I used the feedback when I was on a voice that I got a lot of them correct."

Participant 21 – "It wasn't really helpful to have the different voices because I already knew which voices I was good at and which voices I wasn't good at."

Participant 37 – "I used them especially when I was doing really poorly on one of the trainings. When I was doing really well . . . I didn't use them as much."

3) Did you find yourself developing any specific strategies in terms of what you were listening for?

Participant 5 – "I did, but you had to change what you were listening for in the different voices. For some of the voices I would listen more for the way they would pronounce the sort vowels or long vowels after it. And the other ones I would listen to the breathiness of it."

Participant 8 – "I would always listen for the . . . small pause before the /ħ/ sounds, I would listen to how quickly it descended . . . there was also words that would have the short vowel damma above it, depending on how it sounded in between the letter sound and the damma I could tell which haa it was."

Participant 21 – "Definitely. I think that was the most valuable thing. I learned where in the word the difference is actually heard, because it's not actually heard on the /h/ or /ħ/. It's heard before it or after it. That was the most helpful part about it, I learned where to look for the answer."

Participant 37 – "I think it differed between the voices . . . so I could remember 'Oh yeah, this person pronounces the /ħ/ this way'"

4) Did you find the training beneficial to you as an Arabic learner?

Participant 5 – "It made me a lot more aware . . . of the fact that I need to be differentiating more . . .and I think that it helped me think of a few key differences in the two letters and how they affect the rest of the word. . . Also, the problem with these two letters is actually something that I feel like I can't tackle in my Arabic, I put it on the back burner for so long, so this is actually a really cool thing to do."

Participant 8 – "I had Arabic class today, and I can hear the difference in some words my professor is saying . . . it's just clearer now, it's definitely easier, much more natural . . . it was very useful and it will be useful when I go to Jordan next semester."

Participant 10 – "I really think it was beneficial . . .even when I pronounce it, I have that distinction between the two . . . it helped me to be able to pronounce them correctly."

Participant 21 – "Yes, definitely I think it was good . . .I think that . . . when you get to the advanced levels it's easy to forget about the basics that are really really important. So it was a really good way of reviewing things that are super important if you are going to speak with any level of fluency."

Participant 37 – "I can tell in class now when she says a word . . . I think I am much more confident in being able to say that was a /h/ or a /ħ/ . . . so yeah, I think it was really beneficial in my learning."

These interviews reveal firstly that there was some diversity in how the participants chose to use the website. Some seemed very interested, for example, in tracking their progress over time, while others did not choose to use the progress charts.

Similarly with the feedback, some participants used all of the feedback options, and some only a few, though none of these five indicated that they did not use the feedback at all.

Participant feedback regarding strategies can also be helpful to educators and learners, though no set pattern of strategy use emerges. The comments by Participant 8 suggest that the difficulty of the task may be influenced by the quality of the surrounding vowel. Participant 21 may have found some success listening to the coarticulatory effects of the /ħ/ on surrounding sounds. Furthermore, a few of the participants note that they listened for different things on different voices. This variation in NS pronunciation of /ħ/ may be another reason that learners have so much difficulty perceiving it.

Finally, although a small number, each of the 5 interviewees expressed enthusiasm for the training and considered it a beneficial experience. Participants 8 and 37 in particular made note of specific tangible gains in their perception that they attributed to the training and Participant 10 indicated that he felt that the training had also improved his pronunciation of the /ħ/.

The purpose of the experiment described here was to determine if HVPT could be a useful pedagogical tool for teaching and learning Arabic. Quantitative and qualitative data were reviewed in an effort to determine if the web-based training was indeed effective in improving Arabic learners ability to accurately label /h/ and /ħ/ and to determine if such a training program could be effectively used within the context of a university-level language program in which students have many demands upon their time.

The experiment found statistically significant gains on a perception task for a group of trainees compared to a control group that did not undergo the training. Furthermore, qualitative information collected from the participants that used the training website suggests that most viewed the experience as a positive addition to the Arabic learning, despite the time demands that it placed upon them.

As noted above, some questions remain that are not addressed by the data reviewed and two in particular stand out. First, to what extent, if any, do the gains found in a simple minimal pair identification task carry over to the more real world skills that will determine a student's proficiency in the Arabic language? Data from participants in the interviews suggest that at least in some cases this does in fact happen, though we have no empirical results to rely on at this point. And second, to what extent are the empirically demonstrable improvements in the minimal pair task the result of improved perception and to what extent are they due to improved labeling.

Despite the acknowledged importance of these questions, we believe that sufficient data do exist to warrant the use of phonetic training under certain circumstances. We leave to the final chapter a discussion of how phonetic training can best be used to aid learners of Arabic as well as future avenues of research.

5.6 Conclusion

We have found that an online phonetic training program designed to improve Arabic learners' perception of the /h/-/ħ/ contrast can successfully induce statistically significant improvement in a training group compared to a control group for a lexical identification (minimal pair) task. This result is critical for without it we would have no evidence to cite in promoting the use of phonetic training in Arabic instruction. We are now free to consider how this phonetic training might best be used by teachers and learners of Arabic; however, we will delay this discussion until our final chapter. In the conclusion of this chapter, we will consider ways that this empirical research could be improved upon so that it would yield more reliable results were it to be repeated in the future, or with other consonant contrasts.

Firstly, we must acknowledge that although this research provides evidence that phonetic training can improve the perception of the /h/-/ħ/ contrast, it provides no evidence that high variability phonetic training is the best way to do this as our training group was compared with a control group that underwent no training. More robust results could have been achieved if we had included a second training group that were given low variability training, for example hearing stimuli from only one voice during training. Including this test would be especially helpful in determining if subject improvements were the result merely of improved labeling of the stimuli, or if they actually reflect increased ability to perceive the acoustic differences between /h/ and /ħ/, as discussed in 5.4. Inclusion of a low variability group would also allow this dissertation to contribute more generally to L2 speech perception research and theory.

It would also be of great value, particularly from a pedagogical standpoint, to have included pre- and posttest measures of our participants' ability to pronounce these phonemes. As discussed in Chapter 4 some phonetic training studies have shown evidence for improved pronunciation as a result of perceptual training. If we are able to demonstrate that the same is true for the /h/-/ħ/ contrast it would increase the potential value of the training for our learners.

Finally, the results of this experiment do not give any indication of how long the training gains seen by our participants would last. Although studies of retention have genrally shown that trainees remain significantly better than a control group after several months (see 4.2.1.1) it is not certain that this will be true in all cases. The case for phonetic training as a pedagogical tool is weaker without a test of retention showing that the gains are persistent across time.

Chapter 6: Discussion and Conclusions

6.1 Overview

This dissertation was written in the hopes of improving phonemic perception in native speakers of English learning Arabic in a university setting, and it argues that doing so can be expected to lead to improvements in both listening comprehension and pronunciation. In accordance with a new process approach to listening comprehension a listening subskill (phonetic decoding) was identified that was deemed to present special difficulties for some learners of Arabic. We noted that a significant shortcoming of the process approach to listening comprehension was insufficient knowledge about perceptual processes in the L1 or L2. From there we undertook an in-depth investigation of how native speakers identify the phonemes of their language and also the factors affecting the ability of adult foreign language learners to perceive L2. In an effort to more completely understand the task of our learners we also reviewed research on Arabic phonetic and phonology and the articulatory and acoustic correlates of some of the phonemes that some NS of English struggle to acquire. We identified a laboratory technique in phonetic training that had been developed to test theories of speech perception and second language speech learning and adapted that methodology for a pedagogical context focusing on a very difficult contrast found in Arabic and then tested it on a group of Arabic learners at several different American universities. Our empirical test resulted in significant improvements on a minimal pair test for the trained group compared to a control group with no training, and most importantly, the gains of training transferred to a voice that was not used in the training. Thus we have robust evidence that our learners have made meaningful improvements in their performance of an important listening subskill. Furthermore, this dissertation has brought into existence a website that implements our phonetic training program for at least one contrast and which can serve as a template for training other contrasts in Arabic or in other languages. It also is an excellent starting point for research in other methods for improving segmental perception. We have three aims for this final chapter. First we will review evidence presented earlier in this dissertation in an effort to encapsulate the important principles of second language speech learning that every foreign language teacher should know. This information can help teachers and materials developers be more thoughtful in designing lessons and activities and better able to diagnose listener problems. The second aim of this chapter is to make specific recommendations for the use of phonetic training as a pedagogical tool for Arabic or other languages. We will comment upon the strengths and weaknesses of phonetic training as a pedagogy tool. Finally we will outline a vision both for future research and for the design of an online environment that will serve as a resource for phonetic training for Arabic learners wherever they may be

6.2 Phonetic training as a tool for teaching Arabic

In this section we review the advantages of the phonetic training described in Chapter 5 and explain how we believe that it can best be used.

An important aspect of phonetic training as a teaching tool is that is responds to a stated need in the literature on listening comprehension pedagogy. As described in Chapter 1, the process approach to listening comprehension advocates a more nuanced understanding of the listening process such that the individual subskills of listening can be identified and learners can be given targeted training in those subskills. Furthermore, proponents of this approach have called for a rebalancing of listening pedagogy with a renewed focus on bottom-up processing. The phonetic training program we have created responds to both of these calls. We noted also in Chapter 1 that many advocates of the process approach have insufficient knowledge of the complexities underlying speech perception and that the drills they suggest for improving speech perception are not based on any empirical findings. The phonetic training study we have described here in contrast has been empirically demonstrated to improve learner performance in a language in general way, in at least one perceptual task. For the process approach to achieve its full potential in-depth explorations of perceptual processes, followed by empirical testing of proposed methods is essential and our research has taken an important step in this direction.

Another attractive aspect of phonetic training is that it does not require that teachers devote any of their scarce class time to implement it. This allows them to spend all of that time engaged in more meaningful communicative activities that have been shown to be the best way to build L2 proficiency. This suggests an important "division of labor" in listening pedagogy. Proponents of the process approach are careful to note that more holistic, top-down listening comprehension activities are an important part of a language curriculum. In the early stages of learning it is particularly important that learners receive training in how to listen and how to handle texts that are not 100% comprehensible to them and the classroom is the best place to do this. Students can be explicitly taught how to listen, or can be guided through a process of listening, in the classroom. On the other hand, lower level perceptual skills can, with the help of new technologies, be developed outside of the classroom. Thus classroom listening lessons, like the ones described in Vandergrift & Tafaghodtari (2010), could be used to develop top-down listening skills and to highlight areas of difficulty for individual students who would then have recourse to a number of computer based activities intended to develop the necessary subskills.

A related benefit is that phonetic training can be targeted specifically at those students that need it. As we have seen, there is substantial variation in L2 speech learning outcomes and in the classroom it seems to be the case that some learners are able to quickly learn to perceive phonemes that others struggle with for years. There also may be variation in terms of which phonemes individual learners have difficulty perceiving and ideally it would be possible to incorporate the training of multiple contrasts together with associated diagnostics.

The simplicity of phonetic training further enhances its value as a language-teaching tool. It is not necessary for teachers to learn any new teaching techniques, nor for teachers or learners to acquire any new metalinguistic knowledge about articulators nor their acoustic correlates. Although it is likely to be the case that some aspects of Arabic phonology may require more explicit interventions for at least some learners, it is best to begin with the simplest approach that makes the fewest demands upon scarce resources. Similarly, use of the website does not require any training. The participants in this experiment that used the site all indicated that it was simple and intuitive. Therefore the time spent with the website was all spent on the phonetic training itself.

Finally, phonetic training may be valuable as a tool to help those learners that struggle most with listening comprehension. Research on listening reviewed in the first chapter suggests that it may be their lack of low level processing skills that differentiates bad listeners from good ones. Given the need for strong low level processing to support listening as acquisition as described by Richards (2005), phonetic training may enable hard-working students that might otherwise have fallen behind to keep pace with their colleagues.

6.3 Implementing the Phonetic Training Program

As initially conceived, the phonetic training website was to be used alongside learners' initial contact with Arabic phonology during the first months of a beginning

Arabic class. We now believe that use of phonetic training should not be considered prior to the beginning of the second year of Arabic, or even later. There are two main reasons for this. The first is that, as we have seen, some learners are able to learn to perceive the /ħ/-/h/ contrast through the normal course of interaction with the language and thus those learners' time would be better served engaged in other language learning activities. The second, somewhat related reason is that for a student in the first months of study the 10 hours required by the phonetic training represents a rather large percentage of her total time studying the language. If we estimate that a typical intensive program has 5 contact hours a week with the students, and another 10 hours per week during which the students are working on Arabic outside of class, then after two months of study that student will have spent about 120 hours on Arabic. Thus if phonetic training were part of that mix the student would have spent nearly 10% of her time just learning to discriminate a single Arabic phoneme pair. By the start of the second year however, the student will have spent approximately 480 hours (in and outside of class) on Arabic, and by the end of the second year they will have spent almost 1000 hours with Arabic. If by that time the student has failed to learn to perceive the contrast, phonetic training becomes a viable option.

It is difficult to come to firm conclusions about the right amount of training a learner should undergo. In this experiment, learners were expected to complete 100 training modules representing 2400 minimal pair items with feedback. This number of trials is in line with those of other published phonetic training studies (e.g. Logan et al, 1991; Bradlow et al, 1999; Tajima et al, 2008) . It may be that participants will experience gains with fewer trials, but if 2400 proves insufficient it seems unlikely that further training will be valuable. As we saw in Chapter 4 it appears that the largest gains are made early in training so a learner that has completed 100 training modules without effect is unlikely to benefit from further persistence. A more sophisticated website may be able to accurately predict when a participant's training gains have plateaued, but this is difficult to determine without giving a generalization test; however, it is best to avoid giving an identical generalization test too many times or the participant will become used to it and its results will no longer be reliable. At present, the recommendation is that learners complete the full 100 modules, but no more.

In terms of who should use the training, basically any student that feels they are having difficulty perceiving the difference between /ħ/ and /h/ and that her learning is suffering as a consequence may want to consider phonetic training. Of course, most learners are unlikely to be aware of the phonetic training unless informed by their teachers and so it will be necessary to make teachers aware of the training and its potential benefits. Those teachers in turn may wish to inform students that appear to be struggling. Any student interested in the phonetic training should be made aware that it does require a significant commitment of time, approximately 10 hours over the course of a month, that it is likely to be tedious at times, that progress, even if extant, may be difficult to observe, and that results are not guaranteed. It is important that they know what they are getting into and make an informed decision.

6.4 Limitations of Phonetic Training as a Teaching Tool

The research presented here is intended only as a beginning of an investigation of L2 listening subskills and how they might be improved. This research must be considered in light of its limitations so that educators may make informed decisions about its merits.

A first consideration was noted in Chapter 5, that it is possible that some of the gains uncovered by the training are a reflection of improved labeling rather than improved perception. In other words, the subjects may have already been able to perceive a difference between the two sounds, but did not know which orthographic symbol matched which sound). Improved labeling is of course also of benefit and is of value to the student; however, we wish to claim that we are improving low level perceptual processing abilities in our learners and labeling does not qualify, since it is not a measure of perception. Nevertheless, the ability of phonetic training programs to improve perceptual skills for a number of different L2 phonemes was demonstrated in Chapter 4 and we think that we can confidently claim that some of our participants improved their perceptual skills. Future studies should test perception using discrimination tasks that eliminate the confound of labeling, or should compare one type of training with another such that both groups would be expected to have similar labeling improvement.

Perhaps a more serious concern is this study's lack of content validity. We do not at any time measure our learners listening comprehension skills nor can we cite any empirical evidence that improving learners' ability to perform a minimal pair task causes, or even correlates with, improvements in listening comprehension. We do offer limited qualitative evidence, in the form of our participants' interview responses that at least some of them felt that they had improved their ability to perceive the phonemic contrast in the classroom and one that indicated that he felt he was able to pronounce the /ħ/ better as a result of training. As noted in Chapter 1, in lieu of empirical evidence linking phonetic training with improvements in listening comprehension we have argued that there is an acknowledged need in the field of listening comprehension for tools of this type. At the same time we are aware that it is necessary to continually refine our view of how students learn to listen in a new language, what subskills serve L2 listening, how those subskills can be improved, and how they interact with one another.

A further concern that we have previously hinted at is that the demands that phonetic learning places on students are not trivial. Users of the training system proposed by this study can expect to spend around 10 hours on the training. No amount of creative feedback is likely to change the fact that the training can be very tedious. Learners that spend this time and fail to see results are likely to be very frustrated especially if no further advice is forthcoming on how they might improve their perception. As indicated it is important to manage learner expectations for phonetic training.

6.5 FUTURE DIRECTIONS

6.5.1 Research

The limitations noted above suggest important future directions for researching phonetic training in particular and listening comprehension in general.

Any research that attempts to explicitly link phonetic training, or phoneme perception in general, to proficiency oriented pedagogical outcomes would be welcome. It is exceptionally difficult to measure listening abilities in a principled way and probably impossible to measure the gains that perceptual training might be expected to yield in the short term and so a simple pre- and post-training measure of listening comprehension is unlikely to be informative. It may be possible to measure its benefits to listening comprehension in less direct ways. One method with potential is that found in Field (2003) in which perceptual and contextual cues were in conflict when listeners were asked to identify words in a sentence. As Field notes, this metric needs to be more precise, but it could be an interesting test of the value of phonetic training. It may show for example that learners that have undergone phonetic training are more willing to rely on bottom-up processing. Similarly, carefully designed listening tests, like those in Tsui & Fullilove (1998) in which some questions require more focused, bottom-up listening to answer could be used to detect the influence of training gains.

It would also be worthwhile to explore any potential gains in the pronunciation of the trained phonemes, an original goal of this research that had to be abandoned for methodological reasons. Although prior studies have shown pronunciation gains from perceptual training alone it cannot be assumed that this will be the case for all L2 phonemes and for phonemes like the pharyngeal /ħ/, which is articulatorily novel for our learners, we cannot assume production always follows from perception. Future research should expand upon the phoneme pairs that are trained and take measurements of participant pronunciation gains. If they can reliably be found the case for phonetic training as a language-teaching tool is strengthened.

More generally, research that helps to elucidate the relative contribution of different perceptual subskills – such as phoneme identification and lexical segmentation and access – to overall comprehension and the ways that subskills interact with one another would be most valuable for designing and targeting perceptual training programs. At present it is unclear at what level of perception pedagogical efforts should be focused.

Research exploring the effectiveness of different types of training would also be welcome. The current research compares phonetic training with nothing and finds that it is superior. It would be more valuable to compare one type of training with another and to find what methods yield the best results with the fewest demands on trainees. This research can be guided by the review provided in Chapter 4, and there are three avenues worthy of exploring.

The first is the incorporation of the visual modality in the training stimuli. Recall that Hardison (2003) found significantly greater improvement on an audio-only minimal pair test in perception of English /r/-/l/ by NS of Japanese for participants trained with audiovisual stimuli compared to those with audio stimuli only. Hazan et al (2005) on the other hand found no advantage for incorporating the visual modality in training /r/-/l/ to NS of Japanese, but did find an advantage in training a /b/-/v/-/p/ contrast. Hazan suggests that the visual modality is only important when there is a salient visual cue to the difference. If this is the case, it is unlikely to be helpful in training /ħ/-/h/ or other pharyngeal sounds in Arabic.

A second avenue of exploration that is likely to prove fruitful is the use of adaptive training in which stimuli are initially manipulated to make them easily discriminable and then more and more natural as the learner improves over time. McCandliss et al (2002) compared adaptive training with prototype training and found no advantage for adaptive training. He also found that feedback in the adaptive condition appeared to be unnecessary. McCandliss et al also did not test for generalization to real speech tokens, so the results remain uncertain. The study by Iverson et al (2005) also compared adaptive and prototype training and also found no advantage. They concluded that the ease of implementing prototype training compared to the complexities of creating adaptive stimuli made the former the superior choice. Zhang et al (2009) claim superior benefits for adaptive training of /r/-/l/ to NS of Japanese; however, they are comparing the mean improvement scores from their study to mean improvements scores from other /r/-/l/ training studies and not directly comparing the methods. There is no published phonetic training study that finds a statistically significant advantage for adaptive training, however the studies cited may not have been powerful enough to find differences that do exist and so it remains worth exploring the potential benefits of adaptive training for training Arabic contrasts. It may be particularly beneficial for training learners to perceive the vowel quality differences caused by the secondary emphatic articulation.

A particularly intriguing avenue of inquiry is that of the statistical learning of L2 phoneme perception. Maye (2000) suggests that mere exposure to specially selected distributions of second language sounds may result in at least some phonetic learning. It seems that preattentive perceptual processes are able to track the statistical distribution of speech sounds in the environment and, based on this information, make inferences about the phoneme categories that exist in the language. Research in this area is still very new and no published studies demonstrate statistically based phonetic learning that generalizes to the identification of real speech tokens; however, its potential as a language training tool warrants further investigation. As we noted a major drawback of our own phonetic training program is the demands that it makes on the learners' time and its at times tedious nature. If methods could be found that improve phonetic learning through passive, unattended exposure it would be of great value as it would impose almost no costs upon the learner. Furthermore, new theories of L2 speech learning (e.g. Boersma, Escudero & Hayes, 2003) postulate that initial representations of L2 phonemes are based on statistical distributions of speech sounds and research in this area could help to confirm or deny this proposition.

At present neither adaptive nor statistically based training can be used in training the Arabic /ħ/ contrast because we do not know enough about the acoustic cues relied upon by native speakers in discriminating them and therefore we cannot create stimuli that exaggerate those features, nor can we create distributions of stimuli centered on the relevant acoustic prototypes. There is a clear need for perceptual research with NS of Arabic to determine what cues they rely upon in identifying /ħ/.

The research presented in this dissertation represents an important step towards the development of a new suite of tools for foreign language instruction that are based upon empirical research findings and targeted at the development of specific subskills underlying the L2 listening comprehension process. It represents a promising, though not yet fully explored, avenue of language teaching situated with the new process approach to listening comprehension pedagogy. As noted above, a number of outstanding issues and questions remain and it is not expected that this dissertation will be the last word on the value of phonetic training, nor on its implementation. Next, we lay out our vision for an online learning system designed to improve phonetic perception in learners of Arabic that is based upon the results of this dissertation.

6.5.2 Vision for a comprehensive Arabic phonetic training website

The research presented in this dissertation represents only one small step in a far more ambitious plan to develop an online learning environment for the diagnosis of perceptual difficulties in Arabic learners at all levels and the design of individualized training modules based on this diagnosis. In what follows we lay out our vision for what this learning environment will look like, and then outline the steps that will be necessary to get it there.

As envisioned, the completed learning environment will be designed to diagnose and correct perceptual difficulties in both the perception of individual phonemes, as we have done in the research presented here, and in the perception of phonological quantity, another important and difficult perceptual task that learners of Arabic must master.

Any learner of Arabic that wished to use the site would begin by visiting it and creating an account. At this stage demographic data would be collected such as the user's age, sex, and most importantly, experience with Arabic. It will be especially important at this stage to collect data that will eventually be used to rate the website's efficacy and to identify which potential users are most likely to benefit from the training. In particular, it is necessary to determine how much experience in listening tasks that a given user has prior to using the website. This will help determine if the site is most effective for inexperienced learners or more experienced ones and allow for comparisons between different training methodologies. This will also require that IRB approval be obtained so that users can be asked to opt in to experiments testing the efficacy of different training methodologies. If they are willing to participate in research, participants will be randomly assigned to a control group, in which case they will receive normal prototype based, high variability phonetic training (i.e. the same methodology that was presented in Chapter 5 of this dissertation), or to an experimental group that will receive training by alternate methodologies (for example, adaptive training as described above).

After creating an account, participants will be given a comprehensive diagnostic test designed to assess a battery of different perceptual abilities. As different training modules are added, the diagnostic test will in turn be expanded. Based on the results of this test, participants will be shown the results of their test in comparison with others like them that have also taken the test and be given a report indicating their priorities for phonetic training. Users will be advised as to how they should proceed, but will be allowed to decide for themselves which training modules they wish to use. At present it is deemed best that users complete a full course of training for a given contrast (e.g. /h/-/ħ/) before they begin training on another contrast. This is based on earlier experiments conducted by the present researcher, which attempted to train participants on several contrasts simultaneously. The training gains induced by this method were not statistically significant and participants complained that it placed too many demands upon their time.

At the completion of each training regime, users will be asked to retake the diagnostic test that they took before training so that the gains resulting from the training regime could be assessed. They would also be asked to retake the same test again after three months so that retention of training gains could be measured.

It is already possible to establish the website with the training and testing materials used in the research presented in this dissertation. It requires only that the testing and training websites be merged into a single entity and that functionality for the creation of usernames be implemented. From there the diagnostic test and training modules could slowly be added and the website would slowly grow into its final form. Based upon the research on Arabic presented in Chapter 1, it is recommended that expansion of the website follow the sequence here:

1. The /h/-/ħ/ contrast
2. The emphatic contrast (i.e. /s/-/sˤ/, /t/-/tˤ/, /d/-/dˤ/, /ð/-/ðˤ/)
3. Phonological quantity contrasts (e.g. /i/-/iː/, /s/-/sː/, /fiːsˤaf/-/fiːsˤːaf/)
4. The *hamza-ʕayn* contrast (/ʔ/-/ʕ/)
5. Other phonemic contrasts (/ɣ/-/χ/, /k/-/q/)

As noted, the first item of this list, the /h/-/ħ/ contrast, is ready to be deployed with relatively little effort. The addition of each subsequent contrast would require first the collection of stimuli to be used in the training and testing module. This involves drawing up a list of minimal pairs featuring the contrast in question. Then, speech samples from at least six different native speakers must be collected – five for the training modules and one for the test of generalization. Pilot testing of the samples with other native speakers would provide some certainty that the stimuli are valid examples of the target phonemes. Ideally, each set of training and test stimuli would be tested in an empirical experiment like this one in which the results of training are compared with perceptual changes in an untrained control group. As we have seen though, to do so requires a substantial investment of time and money and may not be feasible for all of the contrast pairs.

Training for phonological quantity contrasts is a little bit more complicated, but the same principles apply. With phonological quantity contrasts it may be more important to vary speaking rate than speaking voices. This is because the assignment of a given phoneme to a long or short category depends on the length of the phoneme relative to other phonemes in the speech stream. A long vowel in rapid speech may be shorter than a short vowel in slow speech. It will definitely be necessary to empirically test the phonological quantity training modules before they are added to the website to be certain that they can indeed do what they claim to do.

6.6 CONCLUSION

The research presented in this dissertation has taken the first important step down a path leading to a new way of diagnosing and treating perceptual difficulties for native speakers of English learning Arabic. We have provided empirical evidence that high variability phonetic training, delivered online and alongside more traditional study, can improve the ability of some learners to perceive the difference between the Arabic laryngeal fricative /h/ and the pharyngeal fricative /ħ/ which has been found to be very difficult for even some advanced learners to acquire. The methodology that we introduce is particularly attractive in that it makes almost no demands on teachers of Arabic – it requires neither that they devote their classroom time to special activities, nor that they acquire any new skills or knowledge about Arabic linguistics. Teachers of Arabic need only be aware of the training website and to make their learners aware of it as well.

Furthermore, the website that we have created represents only the initial stage of a comprehensive online environment for the diagnosis of perceptual problems for learners of Arabic, and for providing training to those learners so that those problems may be overcome. We have also laid out a path for future research and development that can be used to bridge the gap between laboratory research in L2 speech perception and pedagogical practice in the Arabic classroom. It is our sincere hope that the research presented hear can be of value to teachers and learners of other languages as well by providing a blue print for the implementation of phonetic training of all kinds within a more formal language learning curriculum.

Appendix A: Test Documents

RECRUITMENTS SCRIPTS

The text below was circulated to potential participants via their Arabic teachers.

Greetings and Thank you for your interest in my study,

Here is a brief description of what is being asked of you and how the study will work if you choose to participate.

First, to be eligible for the study, you must meet each of the following criteria:

- a student of 2nd or 3rd year Arabic.
- a native speaker of English.
- NOT of Arab heritage.

If you are eligible, and decide to participate, here is how it will work:

1. You will reply to this email, indicating that you are willing to take part in the experiment.
2. I will send you a username. password and url to a website that will test your ability to perceive the difference between هـ and ح.
3. You will visit the website and take the 10-15 minute test. It also includes a brief survey about your language background. You will be paid $5 for taking the pretest.
4. Depending on the outcome of the test and random assignment one of three things will happen:
    a. If you do well on the test, indicating that you have little difficulty perceiving the difference, nothing further will happen.
    b. If you do less well on the exam, indicating room for improvement, you will either be assigned to the control group, or the training group.
        i. If you are assigned to the control group, you will simply retake the online test again after 4 weeks. (and be paid an additional $5)
        ii. If you are assigned to the training group you will:
            - be given a username and password to a second site for phonetic

training.

- agree to complete approximately 10 hours of online phonetic training over a 3-4 week period. The training is online and can be conducted according to your schedule.
- if you complete the training in the required time, and come back for a posttest, you will receive $75 for the training and $5 more for the posttest.

If you have any questions please do not hesitate to contact me at kburnham@utexas.edu or 512-560-2963

Thanks and I look forward to hearing from you.

Kevin Burnham

Department of Middle East Studies

The University of Texas - Austin

Students that indicated a willingness to participate were sent the following email:

Greetings and Thank you for agreeing to participate in my phonetic training experiment. Please carefully read this email before visiting the website.

To take the pretest, please be certain that:

- You are wearing headphones.
- You are using the Firefox browser
- You are using a computer (i.e. not a tablet or smart phone)
- You are in an area with minimal background noise so that you will clearly hear any sounds played over your headphones.
- Your phone is off and you will be able to dedicate your full attention to the test (12-15 minutes)

Please feel free to contact me if you have any further questions at kburnham@utexas.edu or (512) 560-2963

The website is located at: http://kb-pretest.heroku.com/login

Your username is:
Your password is:

Your subject number, which you need for the survey only, is

Please complete the pretest as soon as possible and by Monday September 24 at the latest.

You will be informed of the next step during the week of 9/24.

Thanks and good luck!

Kevin Burnham

CONSENT FORMS

The pretest/training and posttest consent forms are posted here. Consent was obtained online and participants gave consent by clicking 'ok' at the bottom of the consent document.

Pretest consent form

`Title **Phonetic Training for Learners of Arabic** IRB PROTOCOL #
Conducted By:

Primary Investigator	Kevin Burnham (krb287)	Dept. of Middle East Studies	(512) 560-2963
Co-PI/Faculty Sponsor	Mahmoud Al-Batal (mma377)	Dept. of Middle East Studies	(512) 471-3463

You are being asked to participate in a research study. This form provides you with information about the study. The person in charge of this research will also describe this study to you and answer all of your questions. Please read the information below and ask any questions you might have before deciding whether or not to take part. Your participation is entirely voluntary. You can refuse to participate or stop participating at any time without penalty or loss of benefits to which you are otherwise entitled. You can stop your participation at any time and your refusal will not impact current or future relationships with UT Austin or participating sites. To do so simply tell the researcher you wish to stop participation. You may click the link below to save a copy of this form for your records.

The purpose of this study is to explore the potential of a new technique for improving listening comprehension and pronunciation in learners of Arabic. It will include approximately 60 subjects.

If you agree to be in this study, we will ask you to do the following things:

- Take a survey in which you will answer questions about your experience with Arabic and your reasons for studying it.
- Take a computer-based test of your ability to identify certain phonemes of the Arabic language
- Provide your name and email address so that you can be contacted for a follow up test in 4 weeks and so that the results of the two tests can be linked.
- You may be assigned to a training group and be granted access to an online phonetic training system. You may use the system as you choose, but you will be asked to commit to approximately 10 hours of training over 3 weeks. If you are assigned to this group, and complete the required training, you will be given $75 as a token of appreciation for your participation.
- You may be contacted at a later date and asked to participate in a phone interview about your participation in this experiment. You are under no obligation to participate in this interview.

Total estimated time to participate in the pretest is about 20 minutes

Risks of being in the study

- This experiment may involve risks that are currently unforeseeable. If you wish to discuss the information above or any other risks you may experience, you may ask questions now or call the Principal Investigator listed on the front page of this form.

Benefits of being in the study include the possibility that your ability to perceive and pronounce the phonemes of the Arabic language will improve. Furthermore, the experiment has the potential to improve the way that Arabic, and other foreign languages, are taught and learned.

Compensation:

- You will receive $5 for taking this pretest.

Confidentiality and Privacy Protections:

- The data resulting from your participation may be made available to other researchers in the future for research purposes not detailed within this consent form. In these cases, the data will contain no identifying information that could associate you with it, or with your participation in any study.

The records of this study will be stored securely and kept confidential. Authorized persons from The University of Texas at Austin and members of the Institutional Review Board have the legal right to review your research records and will protect the confidentiality of those records to the extent permitted by law. All publications will exclude any information that will make it possible to identify you as a subject. Throughout the study, the researchers will notify you of new information that may become available and that might affect your decision to remain in the study.

Contacts and Questions:

If you have any questions about the study please contact Kevin Burnham. His email and phone are at the top of this form. If you have questions later, want additional information, or wish to withdraw your participation call the researchers conducting the study. Their names, phone numbers, and e-mail addresses are at the top of this page.

If you would like to obtain information about the research study, have questions, concerns, complaints or wish to discuss problems about a research study with someone unaffiliated with the study, please contact the IRB Office at (512) 471-8871. Anonymity, if desired, will be protected to the extent possible. As an alternative method of contact, an email may be sent to orsc@uts.cc.utexas.edu or a letter sent to IRB Administrator, P.O. Box 7426, Mail Code A 3200, Austin, TX 78713.

You will be given a copy of this information to keep for your records.

Statement of Consent:

I have read the above information and have sufficient information to make a decision about participating in this study. I consent to participate in the study.

Click the Button below to consent to participation in the research.

Posttest Consent form

`Title **Phonetic Training for Learners of Arabic** IRB PROTOCOL #
Conducted By:

Primary Investigator	Kevin Burnham (krb287)	Dept. of Middle East Studies	(512) 560-2963
Co-PI/Faculty Sponsor	Mahmoud Al-Batal (mma377)	Dept. of Middle East Studies	(512) 471-3463

You are being asked to participate in a research study. This form provides you with information about the study. The person in charge of this research will also describe this study to you and answer all of your questions. Please read the information below and ask any questions you might have before deciding whether or not to take part. Your participation is entirely voluntary. You can refuse to participate or stop participating at any time without penalty or loss of benefits to which you are otherwise entitled. You can stop your participation at any time and your refusal will not impact current or future relationships with UT Austin or participating sites. To do so simply tell the researcher you wish to stop participation. You may click the link below to save a copy of this form for your records.

The purpose of this study is to explore the potential of a new technique for improving listening comprehension and pronunciation in learners of Arabic. It will include approximately 60 subjects.

If you agree to be in this study, we will ask you to do the following things:

- Take a survey in which you will answer questions about your experience, if any, with the phonetic training website.
- Take a computer-based test of your ability to identify certain phonemes of the Arabic language
- You may be contacted at a later date and asked to participate in a phone interview about your participation in this experiment. You are under no obligation to participate in this interview.

Total estimated time to participate in the pretest is about 20 minutes

Risks of being in the study

- This experiment may involve risks that are currently unforeseeable. If you wish to discuss the information above or any other risks you may experience, you may ask questions now or call the Principal Investigator listed on the front page of this form.

Benefits of being in the study include the possibility that your ability to perceive and pronounce the phonemes of the Arabic language will improve. Furthermore, the experiment has the potential to improve the way that Arabic, and other foreign languages, are taught and learned.

Compensation:

- You will receive $5 for taking this posttest.

Confidentiality and Privacy Protections:

- The data resulting from your participation may be made available to other researchers in the future for research purposes not detailed within this consent form. In these cases, the data will contain no identifying information that could associate you with it, or with your participation in any study.

The records of this study will be stored securely and kept confidential. Authorized persons from The University of Texas at Austin and members of the Institutional Review Board have the legal right to review your research records and will protect the confidentiality of those records to the extent permitted by law. All publications will exclude any information that will make it possible to identify you as a subject. Throughout the study, the researchers will notify you of new information that may become available and that might affect your decision to remain in the study.

Contacts and Questions:

If you have any questions about the study please contact Kevin Burnham. His email and phone are at the top of this form. If you have questions later, want additional information, or wish to withdraw your participation call the researchers conducting the study. Their names, phone numbers, and e-mail addresses are at the top of this page.

If you would like to obtain information about the research study, have questions, concerns, complaints or wish to discuss problems about a research study with someone unaffiliated with the study, please contact the IRB Office at (512) 471-8871. Anonymity, if desired, will be protected to the extent possible. As an alternative method of contact, an email may be sent to orsc@uts.cc.utexas.edu or a letter sent to IRB Administrator, P.O. Box 7426, Mail Code A 3200, Austin, TX 78713.

You will be given a copy of this information to keep for your records.

Statement of Consent:

I have read the above information and have sufficient information to make a decision about participating in this study. I consent to participate in the study.

Click the Button below to consent to participation in the research.

PRETEST SURVEY

The following questions were presented to the participants just prior to the pretest. The survey was administered online using SurveyMonkey.com.

1. What is your subject number?
2. What is your sex?
   a. Male
   b. Female
3. How old are you?
4. Are you a native speaker of English
   a. Yes
   b. No
      i. What is your native language?
5. A 'heritage learner' is defined as "a language student who is raised in a home where a non-English language is spoken, who speaks or at least understands the language, and who is to some degree bilingual in that language and in English." Do you consider yourself to be a heritage learner of Arabic?
   a. No
   b. Yes
      i. What Arabic dialect are you a heritage learner of?
6. A 'heritage learner' is defined as "a language student who is raised in a home where a non-English language is spoken, who speaks or at least understands the language, and who is to some degree bilingual in that language and in English." Do you consider yourself to be a heritage learner of a language OTHER THAN Arabic or English?
   a. Yes

i. What language are you a heritage learner of?

b. No

7. For any country outside the United States that you have lived in for more than three months, please indicate the country, how old you were when you first lived there, and for approximately how long you lived there.
8. For every language that you have studied (not including Arabic) indicate the language and your spoken fluency on a scale of 1(beginner) to 5 (near native).
9. How many SEMESTERS of Arabic study have you completed?
    a. 0-1
    b. 2
    c. 3-4
    d. 5-6
    e. more than 6
10. Indicate you level of agreement with the following statements (Strongly disagree, Disagree, Slightly disagree, Slightly agree, Agree, Strongly agree)
    a. I spend a lot of time studying Arabic.
    b. I believe that it is important to sound like a native speaker when speaking Arabic.
    c. I would like to travel to an Arabic-speaking country.
    d. I feel excited when I hear Arabic spoken.
    e. If I make the effort, I will be able to master the sounds of Arabic.
11. Indicate you level of agreement with the following statements (Strongly disagree, Disagree, Slightly disagree, Slightly agree, Agree, Strongly agree)
    a. The things I want to do in the future require me to learn Arabic.
    b. It is important for me to speak Arabic without an accent.

c. I am willing to put forth a lot of effort to learn Arabic.
d. I like the sounds of Arabic.
e. With enough practice, there are no Arabic words that I cannot pronounce accurately.

12. Indicate you level of agreement with the following statements (Strongly disagree, Disagree, Slightly disagree, Slightly agree, Agree, Strongly agree)
    a. Arabic is important to me because it will help me get a good job.
    b. I like meeting people from Arabic-speaking countries.
    c. I find the differences between English and Arabic vocabulary interesting.
    d. I believe I will be able to speak Arabic without an accent if I keep studying it.
    e. Proper pronunciation is as important as other language skills.
    f. I spend more time on Arabic than on any other subject.
13. Indicate you level of agreement with the following statements (Strongly disagree, Disagree, Slightly disagree, Slightly agree, Agree, Strongly agree)
    a. Compared to my classmates, I study Arabic relatively hard.
    b. Studying Arabic is important to me because I would like to live abroad in an Arabic speaking country.
    c. I am sure I have the necessary skills that enable me to speak accent-free Arabic.
    d. I am interested in the ways that Arabic is used in conversation.
14. Please rate your confidence level in discriminating between هـ and ح where 1 means you have no idea how to tell them apart and 7 means that you never have difficulty telling them apart.

15. Please rate your ability to accurately pronounce the ح where 1 means you can not pronounce it ever and 7 means you pronounce it like a native speaker at all times.
16. You have completed the language background survey. Please close this window or tab and return to the test site. Click on the box “I have completed the survey and am ready for the test” to continue.

POSTTEST SURVEY

The following survey was given to all participants just prior to the posttest. The survey was administered online through SurveyMonkey.com

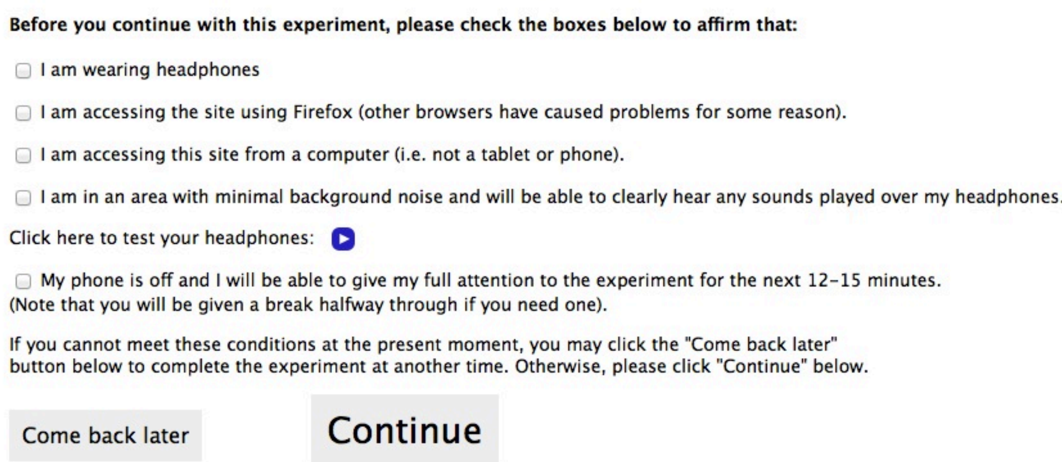
1. What is your subject number?
2. Indicate your extent of agreement with each of the following statements (Strongly disagree, Disagree, Neutral, Agree, Strongly agree).
    a. Overall, I feel that my Arabic has improved since taking the pretest.
    b. I think that my listening comprehension has improved since taking the pretest.
    c. I think that my pronunciation has improved since taking the pretest.
3. Were you assigned to the training group for this experiment?
    a. Yes
    b. No
        i. If 'no' go to 8.
4. Did you complete the required 100 modules?
    a. Yes
    b. No
        i. How many modules did you complete?
            1. 0-50
            2. 51-70
            3. 71-90
            4. 91-99
5. Considering only the ه\ح minimal pairs, indicate the extent to which you agree with each of the following statements. (Strongly disagree, Disagree, Neutral, Agree, Strongly agree) As a result of training:

a. I can now tell a هـ from a ح in isolated words.
b. I can now tell a هـ from a ح when writing out sentences from the DVDs.
c. I can now tell a هـ from a ح during conversations in Arabic.
d. I can now tell a هـ from a ح while doing listening comprehension activities.
e. I can now pronounce a ح when reading individual words.
f. I can now pronounce ح accurately in conversations.

6. Indicate the extent to which you agree with each of the following statements about training (Strongly disagree, Disagree, Neutral, Agree, Strongly agree):
   a. The training was worth the time I spent on it.
   b. I would recommend the website to a friend studying Arabic.
   c. The training should be a required part of the Arabic curriculum.
   d. The training should be an optional part of the Arabic curriculum.
7. Thank you very much for your participation. I am interested in improving the website both in terms of its effectiveness in helping learn Arabic, and its usability and user-friendliness. Would you be willing to participate in a short telephone interview during which you are asked about your experiences with the website? (You don't have to commit to anything at this point.)
   a. Yes
   b. No
8. Thank you very much for your participation. This survey is now complete, please close this window and return to the website.

PRE/POSTTEST SCREEN SHOTS

After being assigned a username and password participants went to http://kb-posttest.heroku.com/login and logged in. They show the consent form (see above) and given a link to the pretest survey. After indicating that they had completed the pretest, participants saw the following:



Before you continue with this experiment, please check the boxes below to affirm that:

- ☐ I am wearing headphones
- ☐ I am accessing the site using Firefox (other browsers have caused problems for some reason).
- ☐ I am accessing this site from a computer (i.e. not a tablet or phone).
- ☐ I am in an area with minimal background noise and will be able to clearly hear any sounds played over my headphones.

Click here to test your headphones:

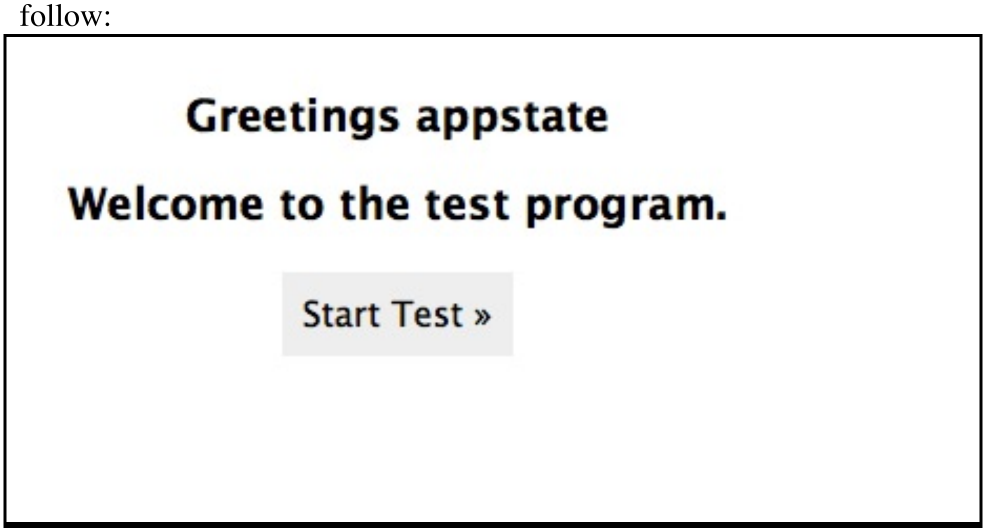
- ☐ My phone is off and I will be able to give my full attention to the experiment for the next 12–15 minutes.

(Note that you will be given a break halfway through if you need one).

If you cannot meet these conditions at the present moment, you may click the "Come back later" button below to complete the experiment at another time. Otherwise, please click "Continue" below.

Come back later Continue

They were given a short 3-item sample test and then told that the main test was to follow:

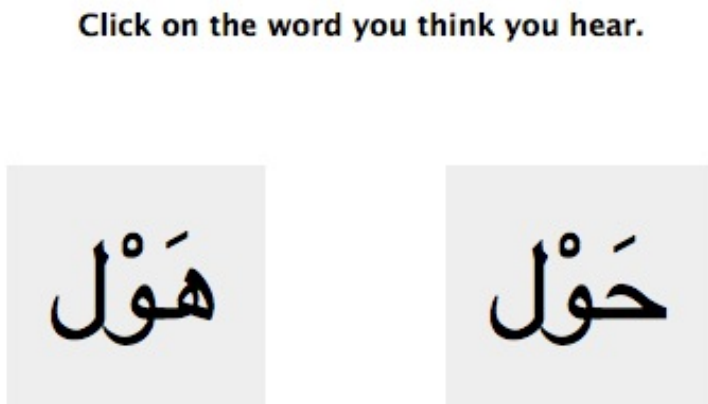


Greetings appstate

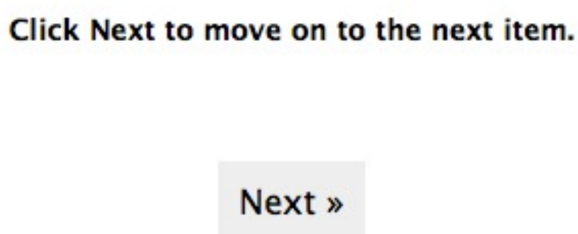
Welcome to the test program.

Start Test »

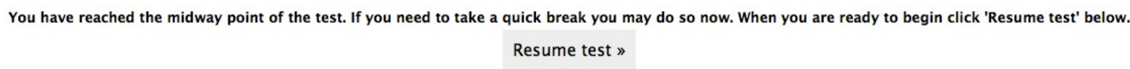
For each test item the user saw two words on the screen, heard one of those words and clicked on the word she thought she heard:



Participants received no feedback on their responses. Between each stimulus they saw this screen:



After completing all items from one voice the participants were given a chance to take a break, after which all stimuli from the other voice were presented.



PRETEST FILES

The table below shows the stimuli used in the minimal pair test given at the beginning and end of the training period. There were two sound files for each of the items below, one from the training voice (TV) and one from the generalization voice (GV).

ItemNo	IPA	Correct Answer	Sound File Name	Distractor	Context
1	/ħubu:b/	حُبوب	H^ubuub	هُبوب	IN
2	/ħawl/	حَوْل	H^awl	هَوْل	IN
3	/ħa:fi/	حافي	H^aafii	هافي	IN
4	/ħuru:b/	حُروب	H^uruub	هُروب	IN
5	/ħari:r/	حَرير	H^ariir	هَرير	IN
6	/harf/	هَرْف	harf	حَرْف	IN
7	/haraqa/	هَرَقَ	haraqa	حَرَقَ	IN
8	/haf:a/	هَفَّ	haffa	حَفَّ	IN
9	/hal:a/	هَلَّ	halla	حَلَّ	IN
10	/ha:mil/	هامِل	haamil	حامِل	IN
11	/balaħ/	بَلَه	balah	بَلَح	FI
12	/jiba:ħ/	جِباه	jibaah	جِباح	FI
13	/ʔaʃba:h/	أشْباح	?a$baaH^	أشْباه	FI
14	/fa:riħ/	فارِح	faariH^	فارِه	FI
15	/taqwi:ħ/	تَقْويح	taqwiiH^	تَقْويه	FI
16	/nabi:ħ/	نَبيح	nabiiH^	نَبيه	FI
17	/naħr/	نَحْر	naH^r	نَهْر	VC
18	/baħr/	بَحْر	baH^r	بَهْر	VC
19	/faħm/	فَحْم	faH^m	فَهْم	VC
20	/ʃahm/	شَهْم	$ahm	شَحْم	VC
21	/ʔahraqa/	أهْرَقَ	?ahraqa	أحْرَقَ	VC
22	/tahdi:d/	تَهْديد	tahdiid	تَحْديد	VC
23	/n:ahiya/	ناهِيَة	naahiya	ناحِيَة	IV
24	/fah:ama/	فَهَّمَ	fahhama	فَحَّمَ	IV
25	/yufaħ:im/	يُفَحِّمُ	yufaH^H^imu	يُفَهِّمُ	IV
26	/faħi:m/	فَحيم	faH^iim	فَهيم	IV
27	/raħab/	رَحَب	raH^ab	رَهَب	IV
28	/hubu:b/	هُبوب	hubuub	حُبوب	IN
29	/hawl/	هَوْل	hawl	حَوْل	IN

30	/ha:fi/	هافي	haafii	حافي	IN
31	/huru:b/	هُروب	huruub	حُروب	IN
32	/hari:r/	هَرير	hariir	حَرير	IN
33	/ħarf/	حَرْف	H^arf	هَرْف	IN
34	/ħaraqa/	حَرَقَ	H^araqa	هَرَقَ	IN
35	/ħaf:a/	حَفَّ	H^affa	هَفَّ	IN
36	/ħal:a/	حَلَّ	H^alla	هَلَّ	IN
37	/ħa:mil/	حامِل	H^aamil	هامِل	IN
38	/balaħ/	بَلَح	balaH^	بَلَه	FI
39	/jiba:ħ/	جِباح	jibaaH^	جِباه	FI
40	/ʔaʃba:h/	أشْباه	?a$baah	أشْباح	FI
41	/fa:rih/	فارِه	faarih	فارِح	FI
42	/taqwi:h/	تَقْويه	taqwiih	تَقْويح	FI
43	/nabi:ħ/	نَبيه	nabiih	نَبيح	FI
44	/nahr/	نَهْر	nahr	نَحْر	VC
45	/bahr/	بَهْر	bahr	بَحْر	VC
46	/fahm/	فَهْم	fahm	فَحْم	VC
47	/ʃahm/	شَحْم	$aH^m	شَهْم	VC
48	/ʔaħraqa/	أحْرَقَ	?aH^raqa	أهْرَقَ	VC
49	/taħdi:d/	تَحْديد	taH^diid	تَهْديد	VC
50	/na:ħiya/	ناحِيَة	naaH^iya	ناهِيَة	IV
51	/faħ:ama/	فَحَّمَ	faH^H^ama	فَهَّمَ	IV
52	/yufah:imu/	يُفَهِّمُ	yufahhimu	يُفَحِّمُ	IV
53	/fahi:m/	فَهيم	fahiim	فَحيم	IV
54	/rahab/	رَهَب	rahab	رَحَب	IV

Appendix B: Training Documents

TRAINING GROUP LETTER

The text below was sent to the participants that were selected for the training group.

I hope that you will be glad to hear that you have been randomly assigned to the **training group** for my phonetic training study. This means that you will be given a username and password to the phonetic training website (see below) and be asked to spend about 10 hours over the next 3-4 weeks using the website to improve your ability to perceive the Arabic ح. At the end of the training period you will retake the same test you just took to measure the improvement in the perception of ح. You will receive $75 for completing the training and $5 for the posttest (in addition to the $5 you will receive for the pretest). Please carefully read these instructions before starting and let me know if you have any questions.

The website is somewhat similar to the pretest that you took. The primary differences are that you are able to hear the word as many times as you would like before giving an answer and that you receive feedback after each answer. Furthermore, the feedback screen gives you the option to hear both versions of the word (with ح and ه) from 5 different voices.

Your username is everything to the left of @ in your email address.
Your password is .

Before you start please note:
- It is best, though not required, to use headphones while using the website.
- You should only use the training from a laptop or desktop computer. Unfortunately the website is not 100% functional on smartphones/tablets.
- Individual modules take from 7-15 minutes each, depending on how you use the available feedback.
- You should spread the required 100 training modules out as much as possible. It is best to do 3-4 modules a day, but the schedule is flexible according to your schedule. I will send you regular reminders in an effort to keep you on schedule.
- If you run into any problems accessing or using the website please let me know immediately at kburnham@utexas.edu or 512-560-2963.
- When you have completed all 100 modules, contact me at kburnham@utexas.edu to arrange for your posttest, which is identical to the pretest that you just took.

To use the website, go to the following url:

http://kb-training.heroku.com/

You may begin using the site immediately.

After logging in, you will see the home screen. On the top left you have the option to see a chart of your progress over time.

If you don't want to see your number correct displayed after each answer, click on 'My Account' at the top right to turn that off.

In the center you can see how many training sessions you have completed, the date of your most recent training, and your best score so far.

Below this box, click 'Start Training' to begin your next training module.

For each test item there are two screens, a test screen, followed by a feedback screen.

On the **test screen** you will see two words written. Above them is a blue button. When you click this button, you will here one of the words spoken. You may listen to the word as many times as you would like. When you are ready, click on the word you think you hear. This will bring you to the feedback screen.

The **feedback** screen will tell you if you were correct or incorrect.

You also have the option on the feedback screen to:

1. Hear the word again, from the same speaker.
2. Hear the other word (the wrong answer) pronounced by the same speaker.
3. Hear either of the two words spoken by 4 different speakers.

When you are ready, click 'Next' to move on to the next item.

Below the 'Next' button you will see your number correct for the current module. Some previous trainees have indicated that they don't like this, so you have the option to turn it off under 'My Account'.

It is best to finish any modules that you start in one session, but if you need to quit early for some reason please be sure to click the 'Exit Training' button. It is far better to exit the training than to speed through it without paying attention.

When you finish all items in a given module, you will be told your score for that module and be returned to the home screen. Click ‘Start Training’ again to start the next module.

Thanks again, I hope that you find the training beneficial to your Arabic studies.

Kevin Burnham

CONTROL GROUP LETTER

This letter was sent to the participants placed in the control group.

Thank you very much for your participation in my experiment. Based on random assignment you have been placed in the control group of my phonetic training experiment. You will not be asked to use the phonetic training website, but I do hope that you will be willing to retake the test you just took in about 4 weeks time. You will be paid an additional $5 for taking this posttest.

I hope that you are not disappointed that you will not yet be able to use the training site, but if it does prove useful in improving perception of the ح , you will be the first to have access to the site if you wish.

I will send you an email in about 3 weeks to ask you to retake the test.

Please send me your mailing address at your earliest convenience so I can send you a check for taking the pretest.

Thanks again and best of luck with you Arabic studies.

bi-t-tawfiiq,

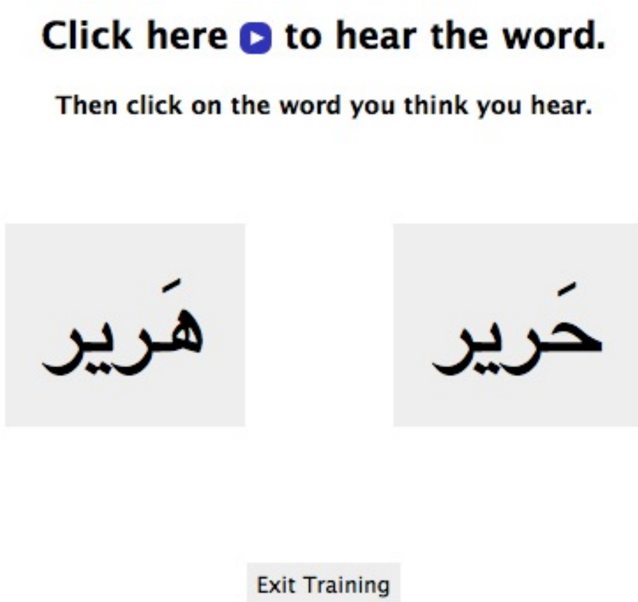
Kevin Burnham

TRAINING WEBSITE SCREEN SHOTS

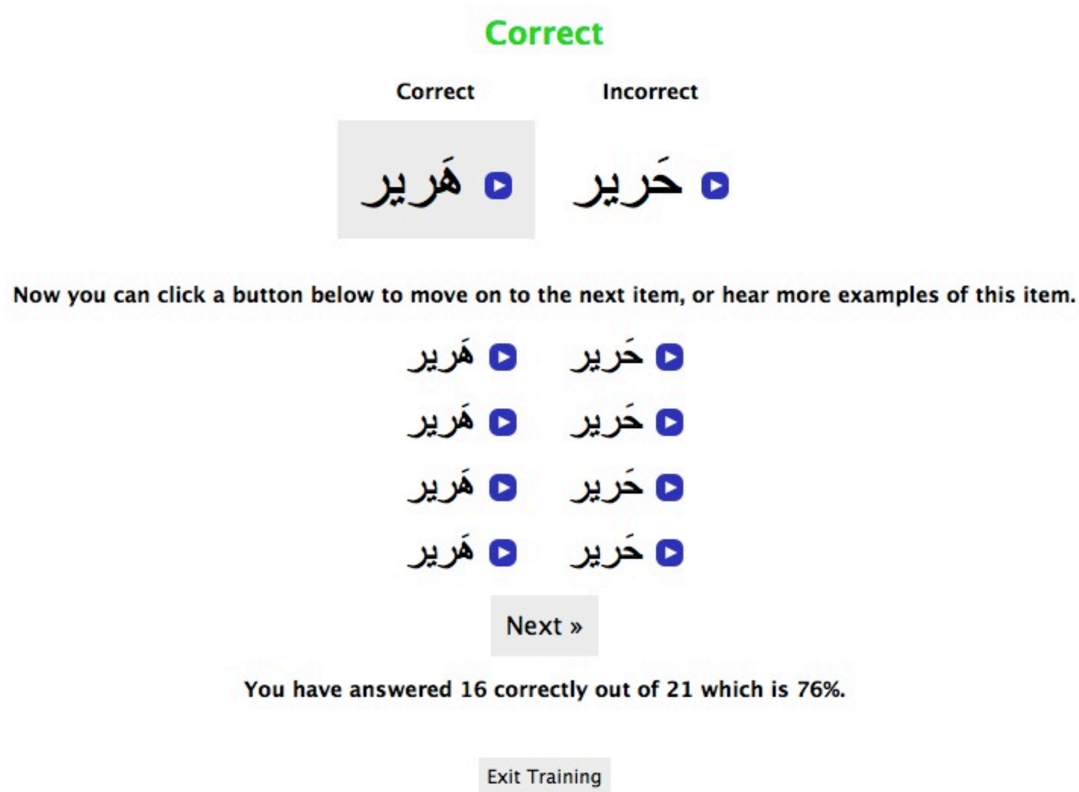
The screen shots below illustrate the functionality of the training website. The homepage is posted below. It indicates the number of sessions that the participant has completed, the date of her last training and the best score she has obtained. A training module begins when the participant clicks on 'Continue Training'.

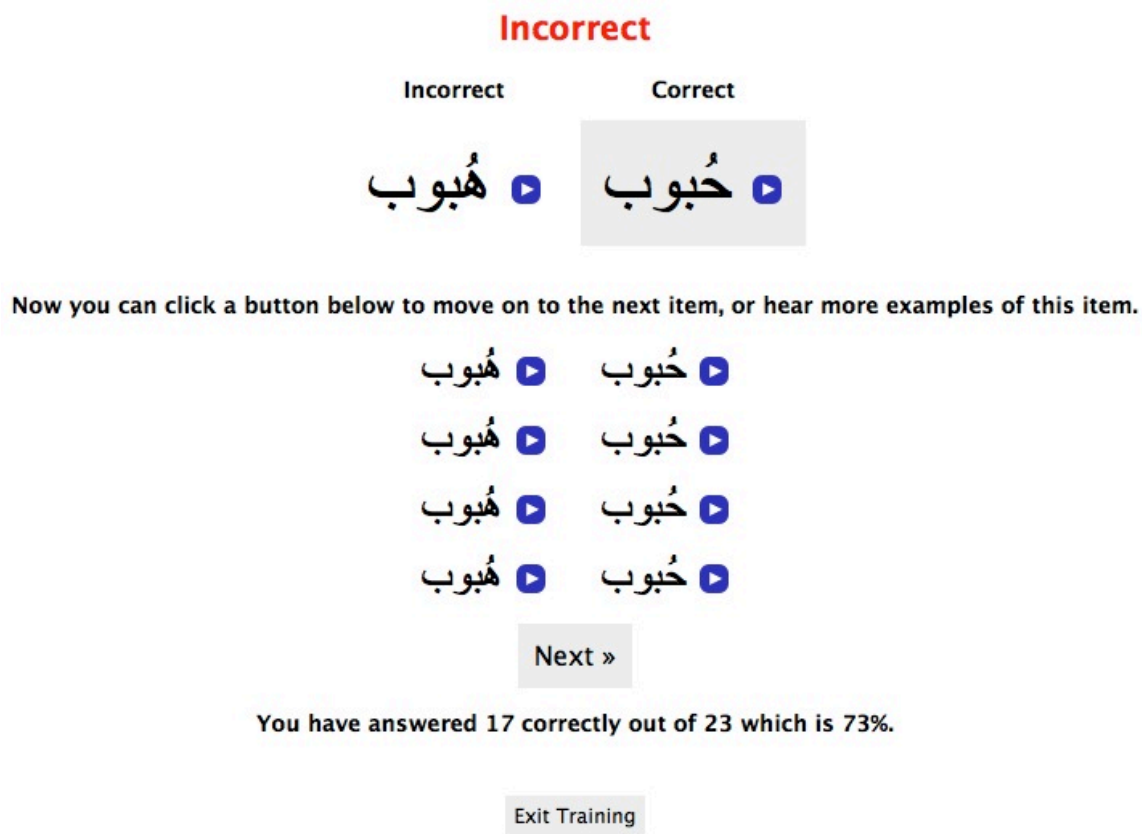
Phoneme contrast	# of Training Sessions completed	Date of last training	Best score obtained
Haahaa pairs	None		

Each training trial consisted of two screens. The test screen, below, gave the participant the opportunity to hear the test word as many times as she liked. When ready, the participant clicked on the word she though she was hearing.



After making a selection the participant would be taken to the feedback screen. This screen indicated to them if they answered correctly or not and gave them the chance to hear the same audio again. They could also hear the distractor word spoken by the same voice, or they could hear either of the two spoken by any of the 4 other voices used in training. When ready, the participant clicked next to begin the next trial. ‘Correct’ and ‘Incorrect’ screens are shown below.





From the home page the participant could access a chart showing their scores over time, by voice.

TRAINING STIMULI

The files used for training trials are indicated in the table below. For each stimulus there were 5 audio files, one for each of the training voices.

ItemNo	IPA	Correct Answer	Sound File Name	Distractor
1	/ħubu:b/	حُبوب	H^ubuub	هُبوب
2	/ħawl/	حَوْل	H^awl	هَوْل
3	/ħa:fi/	حافي	H^aafii	هافي
4	/ħuru:b/	حُروب	H^uruub	هُروب
5	/ħari:r/	حَرير	H^ariir	هَرير
6	/ħarf/	حَرْف	H^arf	هَرْف
7	/haram/	هَرَم	Haram	حَرَم
8	/haraqa/	هَرَقَ	Haraqa	حَرَقَ
9	/haf:a/	هَفَّ	Haffa	حَفَّ
10	/hal:a/	هَلَّ	Halla	حَلَّ
11	/ha:mil/	هامِل	Haamil	حامِل
12	/ha:ra/	هارَ	Haara	حارَ
13	/hubu:b/	هُبوب	Hubuub	حُبوب
14	/hawl/	هَوْل	Hawl	حَوْل
15	/ha:fi/	هافي	Haafii	حافي
16	/huru:b/	هُروب	Huruub	حُروب
17	/hari:r/	هَرير	Hariir	حَرير
18	/harf/	هَرْف	Harf	حَرْف
19	/ħaram/	حَرَم	H^aram	هَرَم
20	/ħaraqa/	حَرَقَ	H^araqa	هَرَقَ
21	/ħaf:a/	حَفَّ	H^affa	هَفَّ
22	/ħal:a/	حَلَّ	H^alla	هَلَّ
23	/ħa:mil/	حامِل	H^aamil	هامِل
24	/ħa:ra/	حارَ	H^aara	هارَ

Appendix C: Interview Documents

INTERVIEW RECRUITMENT SCRIPT

Participants indicating a potential willingness to participate in interviews were sent the following shortly after the end of the training period.

Greetings,

You are receiving this email because you participated in the phonetic training study and indicated a potential willingness to be interviewed about your experiences with the phonetic training website. I am writing now to ask you to participate in such an interview.

If you are willing to participate, I will call you at a prearranged time and ask you a series of questions about your experiences with the website and your opinion of various aspects of the site. The interview will be a maximum of 45 minutes long. Note that your decision as to whether or not to participate in this research will not affect your relationship with your Arabic teacher or the Department of Middle Eastern Studies at UT-Austin. Only I will know which students have participated.

If you are willing to participate, please reply to this email or email me at kburnham@gmail.com and I will send you a document that describes the interview in greater detail.

Thank you sincerely for your time,

Kevin Burnham
UT-Austin

INTERVIEW CONSENT FORM

The consent form below was provided to participants prior to their interview. They were to read it over before the interview. Oral informed consent was obtained over the phone just prior to the interview. Interviews were audio recorded and transcribed.

Title **Phonetic Training for Learners of Arabic** IRB PROTOCOL #
Conducted By:

Primary Investigator	Kevin Burnham (krb287)	Dept. of Middle East Studies	(512) 560-2963
Co-PI/Faculty Sponsor	Mahmoud Al-Batal (mma377)	Dept. of Middle East Studies	(512) 471-3463

You are being asked to participate in a research study. This form provides you with information about the study. The person in charge of this research will also describe this study to you and answer all of your questions. Please read the information you will have the opportunity to ask any questions prior to the start of the study. Your participation is entirely voluntary. You can refuse to participate or stop participating at any time without penalty or loss of benefits to which you are otherwise entitled. You can stop your participation at any time and your refusal will not impact current or future relationships with UT Austin or participating sites. To do so simply tell the researcher you wish to stop participation. The researcher will provide you with a copy of this consent for your records.

The purpose of this study is to explore the potential of a new technique for improving listening comprehension and pronunciation in learners of Arabic. It will include approximately 20 subjects.

If you agree to be in this study, we will ask you to do the following things:

- Participate in a telephone interview about the phonetic training program that you have been using. You will be asked to talk about how valuable you felt the training was, and how it could be improved.

Total estimated time to participate in the interview is a maximum of 45 minutes.

Risks of being in the study

- This experiment may involve risks that are currently unforeseeable. If you wish to discuss the information above or any other risks you may experience, you may ask questions now by calling the Principal Investigator listed on the front page of this form, or you may ask them prior to participation in the interview.

There are no personal benefits to being in the study. However, your participation will potentially aid in the development of a new language teaching methodology that will be of benefit to society at large.

Compensation:

- You will not receive compensation for participation.

Confidentiality and Privacy Protections:

- The data resulting from your participation may be made available to other researchers in the future for research purposes not detailed within this consent form. In these cases, the data will contain no identifying information that could associate you with it, or with your participation in any study.

The records of this study will be stored securely and kept confidential. Authorized persons from The University of Texas at Austin and members of the Institutional Review Board have the legal right to review your research records and will protect the confidentiality of those records to the extent permitted by law. All publications will exclude any information that will make it possible to identify you as a subject. Throughout the study, the researchers will notify you of new information that may become available and that might affect your decision to remain in the study.

Please note that:

- I will audio record the discussion that is about to take place.
- The sound file will be coded so that no information will identify you personally.
- This file will be maintained on a password-protected folder in my laptop.
- Once transcription has occurred, the file will be destroyed.

Contacts and Questions:

For questions about your rights or any dissatisfaction with any part of this study, you can contact, anonymously if you wish, the Institutional Review Board by phone at (512) 471-8871 or email at orsc@uts.cc.utexas.edu.

You will be given a copy of this information to keep for your records.

Statement of Consent:

Prior to your interview, you will be asked to give verbal consent to the study described above.

INTERVIEW QUESTIONS

The text below was used as a guideline for interview questions with training participants.

Introductory script:

Greetings ________________ *and thanks for agreeing to speak with me. First I want to make sure that you understand what will take place before we begin the interview. Have you had a chance to read the consent document that sent you? Do you have any questions about the study or your participation? Do you consent to participation in the study?*

You are encouraged to provide as much information as possible and to be honest in your answering the following questions.

Thank you. I first would like to talk about the functionality of the website. I would like to know if you have any comments about using the website.

Possible follow up questions:
Was the site easy to use? Were there any things about it that you would change?
Did you find the graphs provided to be useful?

Thank you. I am also interested in knowing about how you used the site. For example, how many training modules did you typically complete in one sitting?

Also, did you use the feedback options provided to you? If so, when did you use them? (i.e. after a wrong answer).

In using the website, did you find yourself developing any specific strategies for discriminating between the two phonemes? What were they?

Do you feel that the training was beneficial to you as an Arabic learner? In what ways? (i.e. listening, pronunciation)

Overall, do you think that the time you spent on the training was worthwhile?

What criticisms do you have of the training, how do you think that it might be made more effective?

INTERVIEW TRANSCRIPTS

Do to a technical problem the interview with Participant #10 was not recorded.

Participant 5

In term of the functionality of the website, how did it work for you, was it easy to use, was there anything that you want to change, or just any comments in general.

The website was pretty easy to use – I'm not sure. A few times the responsiveness of it was a little weird, even when I was using the Firefox browser sometimes it would, if my mousse was over to one side and I clicked it, it wouldn't respond. It was [inaudible] I think. Maybe if you could like double click to select, if you have any control over that.

Did you ever use any of the graphs that were provided, the progress charts?

Yeah, I did look at those. I didn't really find them that helpful because the trends didn't really make a lot of sense to me. Like there was a upward trend, but it would go up and down, I don't know if I was like not listening well on certain days or –

So I did like isolate the different voices and stuff like that, that was interesting. And I didn't know if the voices were like trying to exaggerate, like certain parts of how it would sound like [inaudible] or the way it effects short vowels or long vowels. Like different things you would emphasize with the different voices and I don't know if, I couldn't figure out exactly which one went with what, if the chart showed up not like Voice 1, Voice 2, Voice 3 but if it showed like what parts of speech, or what parts of pronunciation you were hearing well, and which parts you weren't, that might have made that more informational, I don't know if that makes sense.

What about the feedback options, did you use those in terms of being able to listen to the different voices hearing the correct and incorrect.

A few times I used the lower [voices] below the current voice, but for the most part I listened to both pronunciations of the same word from the current voice that I was using for that module.

How many training modules would you typically do in one sitting?

It varied a lot. A few times I did between like 5 and 7, I never just did 2 or 3. I did between 5 and 7 most of the time, but otherwise I would do 10 or 12.

Did you find yourself develop any specific strategies in terms of what you were listening for, or how you were making a decision?

I did, but you had to like change what you were listening for the different voices I think. For some of the voices I would be listening more for the way they would pronounce the short vowels or the long vowels after it. And the other ones I would listen to the breathiness of it cause the vowel pronunciation didn't really tell me that much.

But it was different for every voice?

Yes.

Did it seem, in terms of your Arabic worth the time you spent on it?

Yeah, I think it made me a lot more aware or conscientious of the fact that I need to be differentiating more at least. And I think that it helped me think of a few key differences in the two letters and how they affect the rest of the word, at least in the initial part of a word, I picked it up from that point of view. I also, the problem with these two letters is actually something that I feel like I can't tackle in my Arabic, I put it on the back burner for so long, so this is actually a really cool thing to do because [inaudible].

Did you feel that you reached a plateau at any point in the training?

I would say, I think after about 50 of 'em I felt like I understood what the different voices were emphasizing, but I was still making mistakes to the very end so I feel that the repetition was useful.

Are there any specific changes that you would make in terms of how it worked and how it was used or anything about it that annoyed you or would be better differently?

I'm not sure. I think actually that you could design a module so that both versions of the word aren't necessarily there. . . If you clicked on one word and could remember you that far that made it a little bit . . .yeah. Mixing it up maybe would make it more challenging and focus more on the sound. Otherwise yeah, it was cool.

Participant 8

In terms of functionality of the website, do you have any comments about that? Anything you would change, anything that you thought worked well or poorly.

There was nothing that stood out as malfunctioning. In general the website worked very well. But as all things are with technology there are always glitches, Probably throughout the entire course of all 100 lessons, about 20 or so times I would click on the sound icon to play the voice that I was supposed to choose whether it was Haa or haa. It wouldn't play, so I'd have to wait like maybe 30 seconds and then click it again, then it would load the sound. I don't think that's correctable, it's just something I noticed.

In terms of usability was it more or less self-explanatory?

It was very self-explanatory, very intuitive, I couldn't think of a better way to do it honestly.

Did you use the graphs at all, the progress charts that they provided you with?

I used it a lot actually, every 10 or so lessons I would go check how I was doing.

The graphs, were they informative?

They were informative in that I knew which voices to be more careful about when I answered, they were informative in that I learned what . . .cause I would take a guess and I would say okay well maybe I can identify this little thing that might indicate which *haa* sound it was, then I'd play it and, depending on the results over time I would see if it was working or whether it wasn't so I would change strategy depending on how the result [inaudible].

And how many modules would you usually complete in one sitting?

It varied greatly. At the very beginning I did two or three and at the end I was doing 10 or more at a time.

And was that just to get it finished or did you find it more useful to do 10 at a time?

It was mostly cause of midterms I would do a little but just to keep up with it make sure I didn't fall behind, but because it was midterm season when I started it I had to make up for the difference later on.

I would say that probably after about 8 or 9 modules I would get so jaded with the sound that I would stop making sound judgment, so I stopped doing it.

And then, how did you use, if at all, the feedback options . . . did you use those a lot, or did you mostly just go through the tests?

Towards the beginning I did a lot, 'cause I really had no idea how to identify one sound from the other, but then towards the end when I started getting one or two things I could listen for what would regularly correlate with one sound, I didn't need to do it as much.

Did you find yourself developing specific strategies for what you were listening for?

I did, I would always listen for the like *ħaa*, that small pause before the '*h'* sound, I would listen for how quickly the sound descended, I don't really now how to explain it, but I could hear the difference after a while. There was also words that would have the short vowel *damma* above it, depending on how it sounded in between the letter sound and the *damma* I could tell which *haa* it was, whether it dropped into the *damma* or whether it had a kind of an *'a'* sound into the *damma*.

Did you find the training to be beneficial?

Even today, I had Arabic class today, and I can hear the difference in some words that my professor is saying.

And that before was not true?

It was true to some extent, but now it's just clearer now, I wouldn't say that I couldn't do it before, but it's definitely easier, much more natural.

In terms of your overall Arabic learning, do you think it was worth, never mind the compensation, I mean the time you spent on it?

Yeah, absolutely, it's one of those things where I would never do it if I wasn't tested on it. I would never make myself go through and listen to this over and over again. It was very useful, and it will be useful because I am going to Jordan next semester. It will be very useful in trying to determine which sound I'm hearing.

Are there any other specific things about the training? Do you feel like you plateaued at any point?

I'd say I plateaued probably around module 80, if I remember the chart correctly for 2 or 3 voices I was consistently getting better than 80%, and then after 80 I think I only improved on one other voice.

Do you have any other criticisms or thoughts on the training, any ways that it might be better or less tedious or anything like that?

No I think for the amount of information that has to be answered making it anymore complicated would be pointless, it's just in the nature of test to be tedious, I don't think there is anyway around it.

Participant 21

How did [the website] work for you, did you have any comments or suggestions for how it might be improved?

It was a little bit annoying that it didn't work in Chrome, but other than that it wasn't an issue. [inaudible]

Would you say it was easy to use; there wasn't anything confusing about it?

Yeah, definitely.

Did you use the graphs that it had at all to chart your progress?

Yeah.

Were they useful at all, were they helpful were they easy to read?

Yeah, definitely. I think that if it had a kind of weight on like a grade in the class it would be even more helpful. But because it was something I was just doing on the side, I wasn't all that concerned with. You know progress wasn't something that I was like really really concerned with [inaudible]

How many training modules would you typically complete in one sitting?

Usually about 5 in one sitting.

How long would that take?

Two or three minutes.

In terms of the feedback, the different voices that you could listen to, did you use those and in what situations?

You know, I didn't really use them. I generally found when I was answering one of the questions if I got it wrong there were 2 reasons I got it wrong. One was that I wasn't paying attention and I knew that, and the other was that there were some voices, like each voice definitely I could like hear differently, like there were voices that were really hard for me to hear the difference, there were some of them that I was like 'Ok, I'm not going to get that right." Those sound exactly the same to me so whether I got it wrong or right for those I kind of felt it was like 50/50, I'm taking a chance. So it wasn't all that helpful to have the different voices do it because I already knew which voices I was good at and which voices I wasn't good at.

Did you find yourself developing different strategies for what you were listening for when doing the test?

Definitely. I think that was the most valuable thing. I learned where in the word the difference is actually heard cause it's not actually heard on the *haa* or *Haa*. It's heard before it or after it. That was the most helpful part about it, I learned where to look for the answer, if you will.

And so, when you say it was either before or after did that depend on the voice or the word or . . .?

On the word . . . On like how like . . .I felt it was related to how long the word was and what letters were around it. So if it was at the beginning on the word you could hear if it was a *Haa* before the word started. I don't now if you can see this, but I consistently found that the longer the word was it was easier for me.

Did you find overall that it was beneficial? Was it worth the time you spent on it in terms of your Arabic?

Yeah, definitely I think it was good . . I'm in my 3rd year, and I think that like oftentimes in the Arabic classes when you get to the more advanced levels it's easy to forget about the basics that are really really important. So it was a really good way of reviewing things that are super important if your gonna speak with any level of fluency. So it was a good refresher course of how important that is.

Do you feel there was a point where you might have plateaued? At what point do you think it might have leveled out?

I don't think there was any point of that . . I think it would be more beneficial if you did it for a week and then took two weeks off, and then you did it for another week just cause I noticed the more I was doing the better I was doing at it. So on days when I didn't do it at all the next day when I started again I wasn't as good. So I think that might be an interesting way to look at, to think about long term learning in terms of taking breaks, and I don't know, you're the one that's doing a study on this, but like that might be something to look at. The first one back when I hadn't done it for a while was harder. Maybe if you're doing it too often there's a plateau. I found out when I came back to it it was hard again.

Did you notice any carryover into your Arabic class at all, into other Arabic stuff, did it seem to help in that respect at all?

Maybe. Again I think that's a harder thing in the 3rd year position to notice cause in 2nd and 1st year you do a lot of dictation stuff. So being able to hear clues [inaudible] . . .much more useful to me now. I am not dictating things that often. I'm not often writing what I hear, so I'm not sure of any huge difference, but maybe, I don't know.

Participant 37

The first question has to do with the functionality of the website. Did it work well for you, was it easy to use, was there anything about it that you would change?

Yeah, it was actually really easy to use, I used it at a bunch of different computers, and it all worked fine. It was really fast which was awesome, so I really never had any problems with it. It was great.

And in terms of how you interacted with it, is there any specific advice you would have in how to change that?

Not really. I really liked the progress graph, I thought that was really cool, and I looked at that several times, so I though it that was nice. For the most part I thought all of the things that it had, I definitely used them, like I said the progress chart. I thought it was really helpful to have the other voices when you got it right or wrong to see how other people pronounce it. I didn't use it all that much to be honest, but sometimes [inaudible] it was useful to have it as an option.

So in terms of the feedback, you did use the other voices some of the time, but not too frequently?

I used them especially when I was doing really poorly on one of the trainings. I would spend a little bit more time and sometimes that would help me with the next couple of listening exercises. So especially when I was doing badly then I would use them more, but when I was doing really well and it was easy for me to decipher which one it was, I didn't use them as much.

How many modules did you typically complete in one sitting?

I could usually complete anywhere from 4-7. As it got towards the end I usually would do maybe 10 or 12 in a sitting.

And did you think there was some kind of a fall off then in how useful it was after a certain point or it just got too boring?

No I didn't think it got boring, it was just kind of a timing thing, somewhat else I had to do. But I never really tired of it. I thought the length of each one – having 24 – I thought that was a really good amount. Because I didn't feel like it was too long, but at the same time I didn't feel like I heard a substantial amount of the words. I thought by the end I would know the words so well that I would be able to decipher [them] because I had to listened to the voices so many times, cause there were only five, but that wasn't the case either, so I thought 100 was a good amount too.

So you didn't feel that you plateaued in terms of how effective it was?

No, I didn't.

Did you find yourself developing any specific strategies in terms of what you were listening for or how you were trying to make a decision?

A little bit, and I think it differed between the voices because even though I didn't necessarily remember how a specific voice pronounced a certain word, I obviously recognized the voices as I got further into the training, so I could remember after maybe 3 or 4 times listening to that voice during that drill I could remember this person pronounces the harder *'h'* like with more of a Haa like [inaudible], kind of. And so I kind of developed different, like after 3 or 4 I could remember 'Oh I remember this person pronounces the harder *'h'* this way. So that was kind of what I developed to help me do better as I got to maybe 20 or 30 trainings.

And did you find [the training] to be beneficial to you in terms of your Arabic learning?

Yeah I do, I mean I can tell when I'm in class now if she says a word I can almost – I mean I'm not 100% positive 'cause sometimes she won't write the word on the board or something like that -- but I think that I am much more confident in being able to say that was a hard h or that was a soft h. So yeah, I do think that it was really beneficial in my learning.

Would you say then that it was worth the time?

Yeah, absolutely. I really like Arabic, so yeah, definitely I thought it was a great addition to my learning.

Do you have any specific criticisms or things you might change in terms of how it worked or how many modules or anything like that?

No. I think the amount of modules – I mean obviously it is a little bit of a time commitment, but I think you did a good job telling people that in the beginning, so I was not surprised whatsoever, so I thought it was good as long as people know what they are signing up for and they can make the time, because 5 trainings a day would be [inaudible] so no, I really did not have any problem at all with the time commitment or the length of the training.

Bibliography

Abdalla, M. & Al-Batal, M. (2011-2012). College Level Teachers of Arabic in the US: A Survey of Their Professional and Institutional Profiles and Attitudes. *Al-Arabiyya*, 44-45, 1-28.

Abramson, A. & Lisker, L. (1973). Voice-timing perception in Spanish word-initial stops. *Journal of Phonetics*, *1*, 1-8.

Alosh, M. (2009). *Ahlan wa sahlan: Letters and sounds of the Arabic language*. Yale University Press.

Asfoor, M., Ahmad (1982). *Difficulties English speakers encounter in Arabic phonology*. University of San Francisco, United States -- California.

Bagemihl, B. (1995). Language Games and Related Areas. In J. A. Goldsmith (Ed.), *The Handbook of Phonological Theory* (pp. 697-712). Malden MA: Blackwell Publishing.

Best, C., T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171-204). Timonium, MD: York Press.

Best, C., T., McRoberts, G., W., & Sithole, N., M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 345-360.

Best, C., T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics*, *20*(3), 305-330.

Best, C., T., McRoberts, G., W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America*, *109*(2), 775-794.

Best, C. T., Traill, A., Carter, A., Harrison, K. D., & Faber, A. (2003). *!Xóõ click perception by English, Isizulu, and Sesotho listeners*. Proceedings from The 15th International Congress of Phonetic Sciences, Barcelona.

Boersma, P., Escudero, P., & Hayes, R. (2003). *Learning abstract phonological from auditory phonetic categories: an integrated model for the acquisition of language-specific sound categories*. Proceedings from The 15th International Congress of Phonetic Sciences, Barcelona.

Bohn, O.-S., & Flege, J., E.. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics*, *11*(03), 303-328.

Bosch, L., Costa, A., & Sebastián-Gallés, N. (2000). First and second language vowel perception in early bilinguals. *European Journal of Cognitive Psychology*, *12*(2), 189-221.

Bradlow, A., R. (2008). Training non-native language sound patterns: Lessons from training Japanese adults on the English /r/-/l/ contrast. In J. G. Hansen & M. Zampini, L (Eds.), *Phonology and Second Language Acquisition* (pp. 287-308). Philadelphia: John Benjamins Publishing Company.

Bradlow, A., R, Akahane-Yamada, R., Pisoni, D., B, & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics*, *61*(5), 977-985.

Brown, H., Douglas. (1994). *Teaching by principles: An interactive approach to language pedagogy*. Englewood Cliffs, NJ: Prentice Hall Regents.

Brustad, K., Al-Batal, M., & Al-Tonsi, A. (2004). *Alif baa with DVDs: Introduction to Arabic letters and sounds*. Washington, DC: Georgetown University Press.

Brustad K., M. Al-Batal & A. Al-Tonsi (2011). *Al-Kitaab fii Tacallum al-^{c}Arabiyya, Third Edition: Al-Kitaab fii Ta^{c}allum al-^{c}Arabiyya - A Textbook for Beginning Arabic: Part 1*. Georgetown University Press.

Burnham, D., K. (1986). Developmental loss of speech perception: Exposure to and experience with a first language. *Applied Psycholinguistics*, *7*(03), 207-239.

Burnham, K. (2010). *Phonetic discrimination and learners of Arabic*. Middle East Studies Association Annual Meeting, San Diego.

Butcher, A., & Ahmad, K. (1987). Some acoustic and aerodynamic characteristics of pharyngeal consonants in Iraqi Arabic. *Phonetica*, *44*(3), 156-172.

Carney, A., Earley, Widin, G., P., & Viemeister, N., F. (1977). Noncategorical perception of stop consonants differing in VOT. *The Journal of the Acoustical Society of America*, *62*(4), 961-970.

Catford, J. C. (1977). *Fundamental problems in phonetics*. Edinburgh: Edinburgh University Press.

Chang, E., F, Rieger, J., W, Johnson, K., Berger, M., S, Barbaro, N., M, & Knight, R., T. (2010). Categorical speech representation in human superior temporal gyrus. *Nature Neuroscience*, *13*(11), 1428-1432.

Chun, L. S. (2010). Developing intensive listening skills: A case study of the long-term dictation tasks using rapid speech. *The English Teacher*, *XXXIX*, 94-119.

Cooper, F., S., Liberman, A., M., & Borst, J., M. (1951). The interconversion of audible and visible patterns as a basis for research in the perception of speech. *Proceedings of the National Academy of Sciences of the United States of America*, *37*(5), 318-325.

Cross, D., V., & Lane, H., L. (1962). On the discriminative control of concurrent responses: The relations among response frequency, latency, and topography in auditory generalization. *Journal of the Experimental Analysis of Behavior*, *5*(4), 487-496.

Cutting, J. (1982). Plucks and bows are categorically perceived, sometimes. *Attention, Perception, & Psychophysics*, *31*(5), 462-476.

Cutting, J., & Rosner, B. (1974). Categories and boundaries in speech and music. *Attention, Perception, & Psychophysics*, *16*(3), 564-570.

Dehaene-Lambertz, G., Dupoux, E., & Gout, A. (2000). Electrophysiological correlates of phonological processing: A cross-linguistic study. *Journal of Cognitive Neuroscience*, *12*(4), 635-647.

Delattre, P. (1971). Pharyngeal features in the consonants of Arabic, German, Spanish, French, and American English. *Phonetica*, *23*(3), 129-155.

Diehl, R., L., Lotto, A., J., & Holt, L., L. (2004). Speech perception. *Annual Review of Psychology*, *55*(1), 149-179.

Dissoway-Huff, P., Port, R. F., & Pisoni, D. (1982). Context effects in the perception of English /r/ and /l/ by Japanese. In *Research on speech perception, Progress report number 8* (pp. 277-288). Bloomington, IN:.

Dresher, B. E. (2011). The Phoneme. In M. van Oostendorp, C. J. Ewen, E. V. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology* (pp. 241-266). Oxford, England: Wiley-Blackwell.

Ducroquet, L. (1979). The dictation: An outdated exercise? *System*, *7*(2), 125-129.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance*, *25*(6), 1568-1578.

Edman, T., R. (1980). *Learning of intra-phonemic discrimination for several synthetic speech continua*. University of Minnesota, United States -- Minnesota.

Eimas, P., D. (1975). Auditory and phonetic coding of the cues for speech: Discrimination of the [r/l] distinction by young infants. *Perception and Psychophysics*, *18*(5), 341-347.

Elkhafaifi, H. (2005). The effect of prelistening activities on listening comprehension in Arabic learners. *Foreign Language Annals*, *38*(4), 505-513.

Field, J. (1998). Skills and strategies: Towards a new methodology for listening. *ELT Journal*, *52*(2), 110-118.

Field, J. (2000). 'Not waving but drowning': a reply to Tony Ridgway. *ELT Journal*, *54*(2), 186-195.

Field, J. (2002). The changing face of listening. In J. C. Richards & W. A. Renandya (Eds.), *Methodology in language teaching: An anthology of current practice* (pp. 242-247). New York, NY: Cambridge University Press.

Field, J. (2003). Promoting perception: lexical segmentation in L2 listening. *ELT Journal*, *57*(4), 325-334.

Field, J. (2004). An insight into listeners' problems: Too much bottom-up or too much top-down? *System*, *32*(3), 363-377.

Field, J. (2008). *Listening in the language classroom* (Cambridge language teaching library). Cambridge, UK ; New York: Cambridge University Press.

Flege, J. (1987). The production of 'new' and 'similar' phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*, 47-65.

Flege, J., E. (1988). The production and perception of foreign language speech sounds. *Human communication and its disorders: A review*, *2*, 225-401.

Flege, J. (1992). Speech learning in a second language. In C. Ferguson, A., L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 565-604). Timmonium, MD: York Press.

Flege, J., E. (1993). Production and perception of a novel, second-language phonetic contrast. *Journal of the Acoustical Society of America*, *93*, 1589-1589.

Flege, J. (1995a). Second language speech learning: theory, findings and problems. In W. Strange (Ed.), *Speech perception and linguistic experience* (pp. 233-277). Timonium, MD: York Press.

Flege, J., E. (1995b). Two procedures for training a novel second language phonetic contrast. *Applied Psycholinguistics*, *16*(4), 425-442.

Flege, J. (1998). The role of subject and phonetic variables in L2 speech acquisition. In M. Gruber, D. Higgins, K. Olsen & T. Wysocki (Eds.) *Papers from the 34th annual meeting of the Chicago Linguistic Society.*

Flege, J. (2002). Interactions between the native and second-language phonetic systems. In P. Burmeister, T. Piske, & A. Rohde (Eds.), *An integrated view of language development* (pp. 217-243). Trier: Wissenschaftlicher.

Flege, J., E. (1999). Age of learning and second language speech. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 101-131). Mahwah, NJ: Lawrence Erlbaum and Associates.

Flege, J., E, Bohn, O., S, & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. Journal of Phonetics, 25, 437-470.

Flege, J., E., & Hillenbrand, J. (1986). Differential use of temporal cues to the /s/--/z/ contrast by native and non-native speakers of English. *The Journal of the Acoustical Society of America*, *79*(2), 508-517.

Flege, J., E., & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition*, *23*(04), 527-552.

Flege, J., E., MacKay, I., R. A., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *The Journal of the Acoustical Society of America*, *106*(5), 2973-2987.

Flege, J., E, Munro, M., J, & MacKay, I., R.A. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, *97*(5), 3125-3134.

Flege, J., E., Schirru, C., & MacKay, I., R. A. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication*, *40*(4), 467-491.

Flege, J., E., Takagi, N., & Mann, V. (1996). Lexical familiarity and English-language experience affect Japanese adults' perception of /ɹ/ and /l/. *The Journal of the Acoustical Society of America*, *99*(2), 1161-1173.

Flege, J., & Wang, C. (1989). Native-language phonotactic constraints affect how well Chinese subjects perceive the word-final English/t/-/d/contrast. *Journal of Phonetics*, *17*, 299-315.

Flege, J., E., Yeni-Komshian, G., H., & Liu, S. (1999). Age Constraints on Second-Language Acquisition. *Journal of Memory and Language*, *41*(1), 78-104.

Fowler, C., A, & Dekle, D., J. (1991). Listening with eye and hand: Cross-modal contributions to speech perception. *Journal of Experimental Psychology*, *17*(3), 816-828.

Fowler, C. A., & Galantucci, B. (2008). The relation of speech perception and speech production. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 633-652). Malden, MA: Blackwell Publishing.

Francis, A., L., Baldwin, K., & Nusbaum, H., C. (2000). Effects of training on attention to acoustic cues. *Perception & Psychophysics*, *62*(8), 1668-1680.

Francis, A., L., & Nusbaum, H., C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(2), 349-366.

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language*, *47*(1), 27-52.

Ghazeli, S. (1977). *Back consonants and backing coarticulation in Arabic*. The University of Texas at Austin.

Ghowail, T. I. (1987). *The acoustic phonetic study of the two pharyngeals /ħ,ʕ/ and the, two laryngeals /ʔ,h/ in Arabic*. Indiana University, United States -- Indiana.

Gibson, E. J. (1988). Exploratory behavior in the development of perceiving, acting, and the acquiring of knowledge. *Annual review of psychology 39*(1), 1-42.

Goh, C. (2000). A cognitive perspective on language learners' listening comprehension problems. *System*, *28*(1), 55-75.

Goh, C. (2002a). Exploring listening comprehension tactics and their interaction patterns. *System*, *30*(2), 185-206.

Goh, C. (2002b). *Teaching listening in the language classroom*. SEAMEO Regional Language Centre.

Goh, C. (2008). Metacognitive instruction for second language listening development: Theory, practice and research implications. *RELC Journal*, *39*(2), 188-213.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds “l” and “r.”. *Neuropsychologia*, *9*(3), 317-323.

Graham S. 2006. Listening comprehension: The learners’ perspective. *System* 34: 165-182.

Green, K. P., Kuhl, P. K., & Meltzoff, A. N. (1988). Factors affecting the integration of auditory and visual information in speech: The effect of vowel environment. *The Journal of the Acoustical Society of America*, *84*(S1), S155-S155.

Guenther, F., H., Husain, F., T., Cohen, M., A., & Shinn-Cunningham, B., G. (1999). Effects of categorization and discrimination training on auditory perceptual space. *The Journal of the Acoustical Society of America*, *106*(5), 2900-2912.

Guenther, F., H, & Gjaja, M., N. (1996). The perceptual magnet effect as an emergent property of neural map formation. *Journal of the Acoustical Society of America*, *100*(2 Pt 1), 1111-1121.

Guenther, F., H, Nieto-Castanon, A., Tourville, J., A, & Ghosh, S., S (2007). *Effects of Categorization Training on Auditory Perception and Cortical Representations*. Proceedings from the Speech Recognition as Pattern Classification Workshop.

Guion, S., G., Flege, J., E., Akahane-Yamada, R., & Pruitt, J., C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults’ perception of English consonants. *The Journal of the Acoustical Society of America*, *107*(5), 2711-2724.

Hallé, P., A., Best, C., T., & Levitt, A. (1999). Phonetic vs. phonological influences on French listeners’ perception of American English approximants. *Journal of Phonetics*, *27*(3), 281-306.

Hardison, D., M. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics*, *24*(04), 495-522.

Harnad, S. R. (1987). Preface. In S. Harnad, R (Ed.), *Categorical perception: The groundwork of cognition* (pp. ix-x). New York, NY: Cambridge University Press.

Hattori, K., & Iverson, P. (2010). Examination of the Relationship between L2 Perception and Production: An Investigation of English /r/-/l / Perception and Production by Adult Japanese Speakers. In *Second Language Studies: Acquisition, Learning, Education and Technology*.

Hayes, B. (2009). *Introductory phonology*. Malden, MA: Wiley-Blackwell.

Hayes-Harb, R. (2007). Lexical and statistical evidence in the acquisition of second language phonemes. *Second Language Research*, *23*(1), 65-94.

Hazan, V., Sennema, A., Iba, M., & Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication*, *47*(3), 360-378.

Heselwood, B., & Al-Tamimi, F. (2011). A study of the laryngeal and pharyngeal consonants in Jordanian Arabic using nasoendoscopy, videofluoroscopy and spectography. In B. Heselwood & Z. M. Hassan (Eds.), *Instrumental studies in Arabic phonetics* (pp. 101-127). Philadelphia, PA: John Benjamins Publishing Company.

Hockett, C. F. (1955). *A Manual of Phonology*. Baltimore, MD: Waverly Press.

Hoeflaak, A. (2004). Computer-assisted training in the comprehension of authentic french speech: A closer view. *Computer Assisted Language Learning*, *17*(3-4), 315-337.

Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *The Journal of the Acoustical Society of America*, *119*(5), 3059-3071.

Huehnergard, J.(2010). We all write Hieroglyphics: The origin and spread of the alphabet. *The Middle East Outreach Lecture Series*. Lecture conducted from the University of Texas, Austin, TX.

Hulstijn, J. H. (2003). Connectionist models of language processing and the training of listening skills with the aid of multimedia software. *Computer Assisted Language Learning*, *16*(5), 413-425.

Huthaily, K. (2008). *Second language instruction with phonological knowledge: Teaching Arabic to speakers of English*. University of Montana, United States -- Montana.

Højen, A., &Flege, J. (2006). Early learner' discrimination of second-language vowels. *The Journal of the Acoustical Society of America*, *119*, 3072-3084.

Ingram, J., C. L., & Park, S.-G. (1998). Language, context, and speaker effects in the identification and discrimination of English /r/ and /l/ by Japanese and Korean listeners. *The Journal of the Acoustical Society of America*, *103*(2), 1161-1174.

Ingvalson, E., M, McClelland, J., L, & Holt, L., L. (2011). Predicting native English-like performance by native Japanese speakers. *Journal of Phonetics*, *39*(4), 571-584.

IPA Chart (2005), http://www.langsci.ucl.ac.uk/ipa/ipachart.html, available under a Creative Commons Attribution-Sharealike 3.0 Unported License. Copyright © 2005 International Phonetic Association.

Iverson, P., & Evans, B., G. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *The Journal of the Acoustical Society of America*, *126*(2), 866-877.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A. et al. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*(1), B47-B57.

Jamieson, D. G., & Morosan, D. E. (1986). Training non-native speech contrasts in adults: Acquisition of the English /ð/-/θ/ contrast by francophones. *Perception & Psychophysics*, *40*(4), 205-215.

Jamieson, D. G., & Morosan, D. E. (1989). Training new, nonnative speech contrasts: A comparison of the prototype and perceptual fading techniques. *Canadian Journal of Psychology*, *43*(1), 88-96.

Johnson, K. & Mullenix, J. (Eds.). (1997). *Talker Variability in Speech Processing*. San Diego: Academic Press.

Jongman, A., Herd, W., Al-Masri, M., Sereno, J., & Combest, S. (2011). Acoustics and perception of emphasis in Urban Jordanian Arabic. *Journal of Phonetics*, *39*(1), 85-95.

Jusczyk, P. (1977). Perception of syllable-final stop consonants by 2-month-old infants. *Attention, Perception, & Psychophysics*, *21*(5), 450-454.

Jusczyk, P., W. (1997). *The discovery of spoken language*. Cambridge, Mass: MIT Press.

Keating, P. A. (1988). *A survey of phonological features*. Indiana University Linguistics Club.

Keating, P. 2005. *D-prime (signal detection) analysis*. Retrieved July 9, 2013, from http://www.linguistics.ucla.edu/faciliti/facilities/statistics/dprime.htm

Kelly, S., Hirata, Y., Simester, J., Burch, J., Cullings, E., & Demakakos, J. (2008). Effects of hand gesture and lip movements on auditory learning of second language speech sounds. *Acoustical Society of America Journal*, *123*, 3325-3325.

Kluender, K. R., Diehl, R. L., & Killeen, P. R. (1987). Japanese quail can learn phonetic categories. *Science*, *237*(4819), 1195-1197.

Kraus, N., McGee, T., Carrell, T. D., King, C., Tremblay, K., & Nicol, T. (1995). Central auditory system plasticity associated with speech discrimination training. *Journal of Cognitive Neuroscience*, *7*(1), 25.

Kuhl, P., K., & Miller, J., D. (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *The Journal of the Acoustical Society of America*, *63*(3), 905-917.

Kuhl, P., K. (1981). Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech-sound categories. *The Journal of the Acoustical Society of America*, *70*(2), 340-349.

Ladefoged, P. (2005). *Vowels and consonants*. Malden, MA: Blackwell Publishing.

Ladefoged, P., & Broadbent, D., E. (1957). Information conveyed by vowels. *The Journal of the Acoustical Society of America*, *29*(1), 98-104.

Ladefoged, P., & Maddieson, I. (1996). *The Sounds of the world's languages*. Oxford: Blackwell Publishers.

Lane, H. (1965). The motor theory of speech perception: A critical review. *Psychological Review*, *72*(4), 275-309.

Lane, H., L, & Moore, D., J. (1962). Reconditioning a consonant discrimination in an aphasic: an experimental case history. *The Journal of Speech and Hearing Disorders*, *27*, 232-243.

Laufer, A., & Baer, T. (1988). The emphatic and pharyngeal sounds in Hebrew and Arabic. *Language and Speech*, *31*, 181-205.

Lehn, W. (1963). Emphasis in Cairo Arabic. *Language*, *39*(1), 29-39.

Lengeris, A. (2008). *The effectiveness of auditory phonetic training on Greek native speakers' perception and production of Southern British English vowels*.

Proceedings from ISCA Tutorial and Research Workshop on Experimental Psycholinguistics, Athens, Greece.

Liberman, A., M. (1957). Some results of research on speech perception. *The Journal of the Acoustical Society of America*, *29*(1), 117-123.

Liberman, A., M. (1996). *Speech: A Special Code*. Cambridge, MA: MIT Press.

Liberman, A. M., Delattre, P., Cooper, F. S., & Gerstman, L. (1954). The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs: General and Applied*, *68*, 1-13.

Liberman, A. M., Harris, K., S., Kinney, J., Ann, & Lane, H. (1961). The discrimination of relative onset-time of the components of certain speech and nonspeech patterns. *Journal of Experimental Psychology*, *61*(5), 379-388.

Liberman, A., M., Delattre, P., & Cooper, F., S. (1952). The Role of Selected Stimulus-Variables in the Perception of the Unvoiced Stop Consonants. *The American Journal of Psychology*, *65*(4), 497-516.

Liberman, A., M., Harris, K., Safford, Hoffman, H., S., & Griffith, B., C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358-368.

Lisker, L., & Abramson, A., S. (1964). A cross-language study of voicing in initial stops: acoustical measurements. *Word*, *20*, 384-422.

Lisker, L., & Abramson, A. S. (1970). *The voicing dimension: Some experiments in comparative phonetics*. Prague.

Lisker, L. (1986). “Voicing” in English: A Catalogue of Acoustic Features Signaling /b/ Versus /p/ in Trochees. *Language and Speech*, *29*(1), 3-11.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America*, *94*(3), 1242-1255.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *The Journal of the Acoustical Society of America*, *96*(4), 2076-2087.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *The Journal of the Acoustical Society of America*, *89*(2), 874-886.

Logan, J., S., & Pruitt, J., S. (1995). Methodological issues in training listeners to perceive non-native phonemes. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 351-378). Timonium, MD: York Press.

Logan, J., S., Lively, S., E., & Pisoni, D., B. (1993). Training listeners to perceive novel phonetic categories: How do we know what is learned? *The Journal of the Acoustical Society of America*, *94*(2), 1148-1151.

Long, M., H. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition*, *12*(03), 251-285.

Lotto, A. J., Kluender, K. R., & Holt, L. L. (1998). Depolarizing the perceptual magnet effect. *Department of Psychology*.

MacKain, K. S., Best, C., T, & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics*, *2*, 369-390.

MacKay, I., R. A., Flege, J., E., Piske, T., & Schirru, C. (2001). Category restructuring during second-language speech acquisition. *The Journal of the Acoustical Society of America*, *110*(1), 516-528.

Massaro, D., & Cohen, M. (1983). Phonological context in speech perception. *Attention, Perception, & Psychophysics*, *34*(4), 338-348.

Maye, J., C. (2000). *The acquisition of speech sound categories on the basis of distributional information*. The University of Arizona, United States -- Arizona.

Maye, J., & Gerken, L., A (2000). *Learning phonemes without minimal pairs*. Proceedings from 24th Annual Boston University Conference on Language Development, Somerville, MA.

Maye, J., & Gerken, L., A (2001). *Learning Phonemes: How Far Can the Input Take Us?* Proceedings from The 25th Annual Boston University Conference on Language Development, Boston, MA.

Maye, J., Werker, J., F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101-B111.

McCandliss, B. D., Fiez, J. A., Protopas, A., Conway, M., & McClelland, J. L. (2002). Success and failure in teaching the [r]-[l] contrast to Japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective, & Behavioral Neuroscience*, *2*(2), 89-108.

McCarthy, J. J. (1994). The phonetics and phonology of semitic pharyngeals. In P. Keating (Ed.), *Phonological Structure and Phonetic Form: Papers in laboratory phonology III* (pp. 191-233). Cambridge: Cambridge University Press.

McClaskey, C., L., Pisoni, D., B., & Carrell, T., D. (1983). Transfer of training of a new linguistic contrast in voicing. *Perception & Psychophysics*, *34*(4), 323-330.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746-748.

Mehler, J., Bertoncini, J., Barrière, M., & Jassik-Gerschenfeld, D. (1978). Infant recognition of mother’s voice. *Perception* *7*(5), 491-497.

Mendelsohn, D. J. (1994). *Learning to listen: A strategy-based approach for the second-language learner*. San Diego, CA: Dominie Press.

Mendelsohn, D. (2001). Listening comprehension: We've come a long way, but. *Contact*, *27*, 33-40.

Menning, H., Imaizumi, S., Zwitserlood, P., & Pantev, C. (2002). Plasticity of the Human Auditory Cortex Induced by Discrimination Learning of Non-Native, Mora-Timed Contrasts of the Japanese Language. *Learning & Memory*, *9*(5), 267, 253-267, 253.

Miller, J., L. (1994). On the internal structure of phonetic categories: a progress report. *Cognition*, *50*(1-3), 271-285.

Miller, J. L., & Volaitis, L. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics*, *46*(6), 505-512.

Miller, J., D, Wier, C., C, Pastore, R., E, Kelly, W., J, & Dooling, R., J. (1976). Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception. *Journal of the Acoustical Society of America*, *60*(2), 410-417.

Miyawaki, K., Jenkins, J., Strange, W., Liberman, A., Verbrugge, R., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of (r) and (l) by native speakers of Japanese and English. *Perception & Psychophysics*, *18*(5), 331-340.

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition*, *7*(4), 323-331.

Morosan, D., E, & Jamieson, D., G. (1989). Evaluation of a technique for training new speech contrasts: Generalization across voices, but not word-position or task. *Journal of Speech and Hearing Research*, *32*(3), 501-511.

Morris, S. (1983). Dictation: a technique in need of reappraisal. *ELT Journal*, *37*(2), 121-126.

Motohashi-Saigo, Hardison, D., M. (2009). Acquisition of L2 Japanese geminates: Training with waveform displays. *About Language Learning & Technology,* 29-47.

Näätänen, R. (1995). The mismatch negativity: a powerful tool for cognitive neuroscience. *Ear and Hearing*, *16*(1), 6-18.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R., Luuk, A., Allik, J., Sinkkonen, J.. & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, *385*(6615), 432-434.

Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology*, *38*(01), 1-21.

Nishi, K., & Kewley-Port, D. (2007). Training Japanese listeners to perceive American English vowels: Influence of training sets. *Journal of Speech, Language, and Hearing Research*, *50*(6), 1496-1509.

Obrecht, D. H. (1968). *Effects of the second formant on the perception velarization consonants in Arabic*. Mouton.

Oh, J. S., Jun, S.-A., Knightly, L. M., & Au, T. K.-f. (2003). Holding on to childhood language memory. *Cognition*, *86*(3), B53-B64.

Olive, J. P., Greenwood, A., & Coleman, J. (1993). *Acoustics of American English speech: A dynamic approach*. New York, NY: Springer.

Oxford, R. L. (1993). Research update on teaching L2 listening. *System*, *21*(2), 205-211.

Pallier, C., Bosch, L., & Sebastiån-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition*, *64*(3), B9-B17-B9-B17.

Pegg, J., E, & Werker, J., F. (1997). Adult and infant perception of two English phones. *The Journal of the Acoustical Society of America*, *102*(6), 3742-3753.

Peperkamp, S., Michèle, P., & Dupoux, E. (2003). *Allophonic variation and the acquisition of phoneme categories*. Proceedings from the 27th Annual Boston University Conference on Language Development, Sommerville, MA.

Perlman, G. 1990. *dprime - compute d' and beta for signal detection data*. Retrieved July 9, 2013, from http://hcibib.org/perlman/stat/doc/dprime.htm

Peterson, G., E., & Barney, H., L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, *24*(2), 175-184.

Pisoni, D., Bob (1971). *On the nature of categorical perception of speech sounds*. University of Michigan, United States -- Michigan.

Pisoni, D., B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception and Psychophysics*, *13*(2), 253-260.

Pisoni, D. B., Aslin, R. N., Perey, A. J., & Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance*, *8*(2), 297-314.

Pisoni, D., B., & Lazarus, J., House. (1974). Categorical and noncategorical modes of speech perception along the voicing continuum. *The Journal of the Acoustical Society of America*, *55*(2), 328-333.

Pisoni, D., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Attention, Perception, & Psychophysics*, *15*(2), 285-290.

Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *The Journal of the Acoustical Society of America*, *89*(6), 2961-2977.

Polka, L. (1992). Characterizing the influence of native language experience on adult speech perception. *Perception & psychophysics*, *52*(1), 37-52.

Polka, L. (1995). Linguistic influences in adult perception of non-native vowel contrasts. *The Journal of the Acoustical Society of America*, *97*(2), 1286-1296.

Polka, L., Jusczyk, P., W, & Rvachew, S. (1995). Methods for studying speech perception in infants and children. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 49-89). Timonium, MD: York Press.

Polka, L., Colantonio, C., & Sundara, M. (2001). A cross-language comparison of /d /--/ð/ perception: Evidence for a new developmental pattern. *The Journal of the Acoustical Society of America*, *109*(5), 2190-2201.

Port, R. F. (2007). The graphical basis of phones and phonemes. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James E. Flege* (pp. 349-365). Philadelphia, PA: John Benjamins Publishing Company.

Protopapas, A., & Calhoun, B. (2000). *Adaptive phonetic training for second language learners*. Proceedings from InSTIL 2000, Edinburgh, Scotland.

Pruitt, J., S., Jenkins, J., J., & Strange, W. (2006). Training the perception of Hindi dental and retroflex stops by native speakers of American English and Japanese speakers. *The Journal of the Acoustical Society of America*, *119*(3), 1684-1696.

Raphael, L. J. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. *Journal of the Acoustical Society of America 51*(4), 1296-1303.

Raphael, L. J. (2008). Acoustic cues to the perception of segmental phonemes. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 182-206). Alden, MA: Blackwell Publishing, Ltd.

Repp, B. (1981). Two strategies in fricative discrimination. *Attention, Perception, & Psychophysics*, *30*(3), 217-227.

Repp, B., H. (1984). Categorical perception: Issues, methods, findings. In J. Lass (Ed.), *Speech and Language, Vol.10* (pp. 243-335). New York: Academic Press.

Richards, J., C. (1983). Listening comprehension: Approach, design, procedure. *TESOL Quarterly*, *17*(2), 219-240.

Richards, J. C. (2005). Second thoughts on teaching listening. *RELC Journal*, *36*(1), 85-92.

Ridgway, T. (2000). Listening strategies - I beg your pardon? *ELT Journal*, *54*(2), 179-185.

Rochet, B., L. (1995). Perception and production of second-language speech sounds by adults. *Speech perception and linguistic experience: Issues in cross-language research*, 379-410.

Rosen, S., & Howell, P. (1987). Auditory, articulatory and learning explanations of categorical perception in speech. In S. Harnad, R (Ed.), *Categorical perception: The groundwork of cognition* (pp. 113-160). New York, NY: Cambridge University Press.

Rosenblum, L. D. (2008). Primacy of multimodal speech perception. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 51-78). Malden, MA: Blackwell Publishing.

Samuel, A. G. (1977). The effect of discrimination training on speech perception: Noncategorical perception. *Perception and Psychophysics*, *22*(4), 321-330.

Sebastián-Gallés, N., & Soto-Franco, S. (1999). Online processing of native and non-native phonemic contrasts in early bilinguals. *Cognition*, *72*(2), 111-123.

Shahin, K. (2011). A phonetic study of guttural laryngeals in Palestinian Arabic using laryngoscopic and acoustic analysis. In B. Heselwood & Z. M. Hassan (Eds.), *Instrumental Studies in Arabic Phonetics* (pp. 129-140). Philadelphia, PA: John Benjamins Publishing Company.

Sharma, A., & Dorman, M., F. (2000). Neurophysiologic correlates of cross-language phonetic perception. *The Journal of the Acoustical Society of America*, *107*(5), 2697-2703.

Sheerin, S. (1987). Listening comprehension: teaching or testing? *ELT Journal*, *41*(2), 126-131.

Sheldon, A., & Strange, W. (1982). The Acquisition of /r/ and /l/ by Japanese Learners of English: Evidence That Speech Production Can Precede Speech Perception. *Applied Psycholinguistics*, *3*(3), 243-261.

Sibawayh. (1988). *Al-Kitaab*. Cairo: Maktabat Al-Khaniji.

Sinnott, J., M., Gonzales, C., L., Masood, A., F., & Ishihara, T. (2007). Training humans in non-native phoneme perception using a monkey psychoacoustic procedure. *The Journal of the Acoustical Society of America*, *121*(6), 3846-3857.

Stansfield, C. W. (1985). A history of dictation in foreign language teaching and testing. *The Modern Language Journal*, *69*(2), 121-128.

Strange, W. (1972). *The effects of training on the perception of synthetic speech sounds: Voice onset time*. University of Minnesota, United States -- Minnesota.

Strange, W. (2007). Cross-language similarity of vowels: Theoretical and methodological issues. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning* (pp. 35-55). Philadelphia, PA: John Benjamins Publishing Company.

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r/-/l/ by Japanese adults learning English. *Perception & Psychophysics*, *36*, 131-145.

Streeter, L., A. (1976). Language perception of 2-mo-old infants shows effects of both innate mechanisms and experience. *Nature*, *259*(5538), 39-41.

Studdert-Kennedy, M., Liberman, A., M., Harris, K., S., & Cooper, F., S. (1970). Motor theory of speech perception: A reply to Lane's critical review. *Psychological Review*, *77*(3), 234-249.

Sussman, H., M. (1986). A neuronal model of vowel normalization and representation. *Brain and Language*, *28*(1), 12-23.

Tajima, K., Rothwell, A., & Munhall, K., G. (2002). Native and non-native perception of phonemic length contrasts in Japanese: Effect of identification training and exposure. *The Journal of the Acoustical Society of America*, *112*(5), 2387-2387.

Tajima, K., Kato, H., Rothwell, A., Akahane-Yamada, R., & Munhall, K., G. (2008). Training English listeners to perceive phonemic length contrasts in Japanese. *The Journal of the Acoustical Society of America*, *123*(1), 397-413.

Takagi, N. (2002). The limits of training Japanese listeners to identify English /r/ and /l/: Eight case studies. *The Journal of the Acoustical Society of America*, *111*(6), 2887-2896.

Tees, R. C., & Werker, J. F. (1984). Perceptual flexibility: maintenance or recovery of the ability to discriminate non-native speech sounds. *Canadian Journal of Psychology*, *38*(4), 579-590.

Tremblay, K., Kraus, N., Carrell, T. D., & McGee, T. (1997). Central auditory system plasticity: Generalization to novel stimuli following listening training. *The Journal of the Acoustical Society of America*, *102*(6), 3762-3773.

Tsui, A. B. M., & Fullilove, J. (1998). Bottom-up or top-down processing as a discriminator of L2 listening performance. *Applied Linguistics*, *19*(4), 432-451.

Tsushima, T., Takizawa, O., Sasaki, M., Shiraki, S., Nishi, K., Kohno, M. et al. (1994). *Discrimination of English/r/-/l/and/w-y/by Japanese infants at 6-12 months: Language-specific developmental changes in speech perception abilities*. Proceedings from Third International Conference on Spoken Language Processing, Yokohama, Japan.

Tyler, M. D. (2001). Resource consumption as a function of topic knowledge in nonnative and native comprehension. *Language Learning*, *51*(2), 257-280.

Underbakke, M., Polka, L., Gottfried, T., L., & Strange, W. (1988). Trading relations in the perception of /r/--/l/ by Japanese learners of English. *The Journal of the Acoustical Society of America*, *84*(1), 90-100.

Vandergrift, L., & Tafaghodtari, M. H. (2010). Teaching L2 learners how to listen does make a difference: An empirical study. *Language Learning*, *60*(2), 470-497.

Ventureyra, V. A. G., Pallier, C., & Yoo, H.-Y. (2004). The loss of first language phonetic perception in adopted Koreans. *Journal of Neurolinguistics*, *17*(1), 79-91.

Volaitis, L., E., & Miller, J., L. (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *The Journal of the Acoustical Society of America*, *92*(2), 723-735.

Walley, A. C. (2007). Speech perception in childhood. In D. B. Pisoni & R. E. Remez (Eds.), *The Handbook of Speech Perception* (pp. 449-468). Malden, MA: Blackwell Publishing.

Wang, Y., Jongman, A., & Sereno, J., A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *The Journal of the Acoustical Society of America*, *113*(2), 1033-1043.

Wang, X., & Munro, M., J. (2004). Computer-based training for learning English vowel contrasts. *System*, *32*(4), 539-552.

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America*, *106*(6), 3649-3658.

Wayland, R., P., & Miller, J. L. (1992). *Influence of internal phonetic category structure in on-line speech processing*. Proceedings from 33rd Annual Meeting of the Psychonomic Society, St. Louis, MO.

Wayland, R., P., & Guion, S., G. (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning*, *54*(4), 681-712.

Wehr, H. (1994). *A dictionary of modern written Arabic: (Arabic-English)*. Ithaca, NY: Spoken Language Services.

Werker, J., F, & Tees, R., C. (1983). Developmental changes across childhood in the perception of non-native speech sounds. *Canadian Journal of Psychology*, *37*(2), 278-286.

Werker, J., F, & Lalonde, C., E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental psychology*, *24*(5), 672-683.

Werker, J., F., Gilbert, J., H. V., Humphrey, K., & Tees, R., C. (1981). Developmental aspects of cross-language speech perception. *Child Development*, *52*(1), 349-355.

Werker, J., F., & Tees, R., C. (1984a). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49-63.

Werker, J. F., & Tees, R. C. (1984b). Phonemic and phonetic factors in adult cross-language speech perception. *The Journal of the Acoustical Society of America*, *75*(6), 1866-1878.

Wilson, M. (2003). Discovery listening - improving perceptual processing. *ELT Journal*, *57*(4), 335-343.

Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., Czigler, I., Csépe, V., Iimoniemi, R., Näätänen, R. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology*, *36*(05), 638-642.

Yamada, R., A. (1995). Age and acquisition of second language speech sounds: Perception of American English /r/ and /l/ by native speakers of Japanese. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 305-320). Timonium, MD: York Press.

Yeou, M. (1995). *Trading relations between cues for the pharyngealized/non-pharyngealized contrast*. Proceedings from 13th ICPhSc, Stockholm, Sweden.

Yeou, M., & Maeda, S. (2011). Airflow and acoustic modeling of pharyngeal and uvular consonants in Moroccan Arabic. In B. Heselwood & Z. M. Hassan (Eds.), *Instrumental studies in Arabic phonetics* (pp. 141-162). Philadelphia, PA: John Benjamins Publishing Company.

Vita

Kevin Burnham graduated from Hamden High School in Hamden, CT in 1990. In the same year he entered Bucknell University in Lewisburg, PA from where he received a Bachelor of Arts in Russian in 1995. In 2003 he received a Masters of Arts in Economics from Portland State University and in 2006 a Masters of Science in Arabic Language, Literature and Linguistics from Georgetown University.

Permanent address (or email): kburnham@gmail.com

This dissertation was typed by the author